

Faster Query Answering in Probabilistic Databases using Read-Once Functions*

Sudeepa Roy[†]Vittorio Perduca[‡]Val Tannen[§]

Abstract

A boolean expression is in *read-once form* if each of its variables appears exactly once. When the variables denote independent events in a probability space, the probability of the event denoted by the whole expression in read-once form can be computed in polynomial time (whereas the general problem for arbitrary expressions is #P-complete). Known approaches to checking read-once property seem to require putting these expressions in disjunctive normal form. In this paper, we tell a better story for a large subclass of boolean event expressions: those that are generated by conjunctive queries without self-joins and on tuple-independent probabilistic databases.

We first show that given a tuple-independent representation and the *provenance graph* of an SPJ query plan without self-joins, we can, without using the DNF of a result event expression, efficiently compute its *co-occurrence graph*. From this, the read-once form can already, if it exists, be computed efficiently using existing techniques. Our second and key contribution is a complete, efficient, and simple to implement algorithm for computing the read-once forms (whenever they exist) directly, using a new concept, that of *co-table graph*, which can be significantly smaller than the co-occurrence graph.

*A shorter version of this paper will appear in the proceedings of ICDT 2011.

[†]University of Pennsylvania. Email: sudeepa@cis.upenn.edu. Supported by NSF Award IIS-0803524.

[‡]University of Pennsylvania. Email: perduca@cis.upenn.edu. Supported by NSF Award IIS-0629846.

[§]University of Pennsylvania. Email: val@cis.upenn.edu. Supported in part by NSF Awards IIS-0629846 and IIS-0803524.

$$\begin{array}{c}
R = \begin{array}{|c|} \hline a_1 \\ \hline b_1 \\ \hline a_2 \\ \hline \end{array} \begin{array}{l} 0.3 \\ 0.4 \\ 0.6 \end{array} \quad
S = \begin{array}{|c|c|} \hline a_1 & c_1 \\ \hline b_1 & c_1 \\ \hline a_2 & c_2 \\ \hline a_2 & d_2 \\ \hline \end{array} \begin{array}{l} 0.1 \\ 0.5 \\ 0.2 \\ 0.1 \end{array} \quad
T = \begin{array}{|c|} \hline c_1 \\ \hline c_2 \\ \hline d_2 \\ \hline \end{array} \begin{array}{l} 0.7 \\ 0.8 \\ 0.4 \end{array}
\end{array}
\quad (a)$$

$$\begin{array}{c}
R = \begin{array}{|c|} \hline a_1 \\ \hline b_1 \\ \hline a_2 \\ \hline \end{array} \begin{array}{l} w_1 \\ w_2 \\ w_3 \end{array} \quad
S = \begin{array}{|c|c|} \hline a_1 & c_1 \\ \hline b_1 & c_1 \\ \hline a_2 & c_2 \\ \hline a_2 & d_2 \\ \hline \end{array} \begin{array}{l} v_1 \\ v_2 \\ v_3 \\ v_4 \end{array} \quad
T = \begin{array}{|c|} \hline c_1 \\ \hline c_2 \\ \hline d_2 \\ \hline \end{array} \begin{array}{l} u_1 \\ u_2 \\ u_3 \end{array}
\end{array}
\quad (b)$$

$$Q() : \neg R(x), S(x, y), T(y) \quad (c)$$

Figure 1: (a) A tuple-independent probabilistic database. (b) Event table representation. (c) An *unsafe* query.

1 Introduction

The computation of distributions for query answers on probabilistic databases is closely related to the manipulation of boolean formulas. This connection has led both to interesting theoretical questions and to implementation opportunities. In this paper we consider the *tuple-independent* model¹ for probabilistic databases. In such a model probabilistic databases are represented by tables whose tuples t are each annotated by a probability value $p_t > 0$, see Fig. 1(a). Each tuple appears in a possible world (instance) of the representation with probability p_t independently of the other tuples. This defines a probability distribution on all possible instances.

Manipulating all possible instances is impossibly unwieldy so techniques have been developed [27, 11, 36] for obtaining the query answers from the much smaller representation tables. This is where boolean formulas make their entrance. The idea, by now well-understood [8, 2, 10, 1], is to define the relational algebra operators on tables whose tuples are annotated with event expressions. The event expressions are boolean expressions whose variables annotate the tuples in the input tables. The computation of event expressions is the same as that used in c-tables [22] as models for incomplete and probabilistic databases are closely related [16]. Once the event expressions are computed for the tuples in the representation table of the query answer (which is in general *not* tuple-independent), probabilities are computed according to the standard laws.

The event expressions method was called *intensional* “semantics” by Fuhr and Rölleke and they observed that with this method computing the query answer probabilities seems to require exponentially many steps in general [11]. Indeed, the data complexity² of query evaluation on probabilistic databases is #P-complete, even for conjunctive queries [13], in fact even for quite

¹This model has been considered as early as [4] as well as in, eg., [11, 36, 13, 8].

²Here, and throughout the paper, the data input consists of the representation tables [13, 8] rather than the collection of possible worlds.

simple boolean queries [8] such as the query in Fig. 1(c).

But Fuhr and Rölleke also observe that certain event independences can be taken advantage of, when present, to compute answer probabilities in PTIME, with a procedure called *extensional* “semantics”. The idea behind the extensional approach is the starting point for the remarkable results [8, 9] of Dalvi and Suciu who discovered that the conjunctive queries can be decidably and elegantly separated into those whose data complexity is #P-complete and those for whom a *safe plan* taking the extensional approach can be found.

Our starting point is the observation that even when the data complexity of a query is #P-complete (i.e. the query is *unsafe* [8]), there may be classes of data inputs for which the computation can be done with the extensional approach, and is therefore in PTIME. We illustrate with a simple example.

Example 1. Consider the tuple-independent probabilistic database and the conjunctive query Q in Fig. 1. Since the query Q is boolean it has just one possible answer and the event expression annotating it is ³

$$f = w_1v_1u_1 + w_2v_2u_1 + w_3v_3u_2 + w_4v_4u_3 \quad (1)$$

This was obtained with the standard plan $\pi_{\emptyset}((R \bowtie S) \bowtie T)$. However, it is equivalent to another boolean expression

$$(w_1v_1 + w_2v_2)u_1 + w_3(v_3u_2 + v_4u_3) \quad (2)$$

which has the property that each variable occurs exactly once.

Event (boolean) expressions in which each variable occurs exactly once are in *read-once form* (see [28]). For read-once forms the events denoted by non-overlapping subexpressions are jointly independent, so we can use the key idea of the extensional approach:

Fact If events E_1, \dots, E_n are jointly independent then

$$P(E_1 \cap \dots \cap E_n) = P(E_1) \dots P(E_n) \quad (3)$$

$$P(E_1 \cup \dots \cup E_n) = 1 - [1 - P(E_1)] \dots [1 - P(E_n)]. \quad (4)$$

Example 2 (Example 1 continued). The probability of the answer (2) can be computed as follows

$$\begin{aligned} P(f) &= P((w_1v_1 + w_2v_2)u_1 + w_3(v_3u_2 + v_4u_3)) = \\ &= 1 - [1 - P(w_1v_1 + w_2v_2)P(u_1)][1 - P(w_3)P(v_3u_2 + v_4u_3)] \end{aligned}$$

where

$$P(w_1v_1 + w_2v_2) = 1 - [1 - P(w_1)P(v_1)][1 - P(w_2)P(v_2)]$$

and

$$P(v_3u_2 + v_4u_3) = 1 - [1 - P(v_3)P(u_2)][1 - P(v_4)P(u_3)].$$

We can extend this example to an entire class of representation tables of unbounded size. For each n , the relations R, T will have $3n$ tuples while S will have $4n$ tuples, and the probability of the answer can be computed in time $O(n)$, see Appendix A.

It is also clear that there is no relational algebra plan that directly yields (2) above.

³To reduce the size of expressions and following established tradition we use $+$ for \vee and \cdot for \wedge , and we even omit the latter in most terms.

In fact, Fuhr and Rölleke (see [11], Thm.4.5) state that probabilities can be computed by “simple evaluation” (i.e., by the extensional method) if and only if the event expressions computed intensionally are in read-once form. Moreover, the *safe plans* of [8] are such that all the event expressions computed with the intensional method both on intermediary relations and on the final answer, are in read-once form.

But more can be done. Notice that the expression (1) is not in read-once form but it is equivalent to (2) which is. Boolean expressions that are equivalent to read-once forms have been called by various names, eg., separable, fanout-free [20], repetition-free [17], μ -expressions [35], non-repeating [32], but since the late 80’s [21] the terminology seems to have converged on *read-once*. Of course, not all boolean expressions are read-once, eg., $xy + yz + zx$ or $xy + yz + zu$ are not.

With this motivation we take the study of the following problem as the goal of this paper:

Problem Given tuple-independent database I and boolean conjunctive query Q , when is $Q(I)$ read-once and if so, can its read-once form be computed efficiently?

It turns out that [12] gives a fast algorithm that takes a formula in irredundant disjunctive normal form, decides whether it is read-once, and if it is, computes the read-once form (which is in fact unique modulo associativity and commutativity). The algorithm is based upon a characterization in terms of the formula’s co-occurrence graph given in [18].

Some terminology (taken up again in Section 2). Since we don’t have anything to say about negation or difference in queries we work only with *monotone* boolean formulas (all literals are positive, only disjunction and conjunction operations). Disjunctive normal forms (DNFs) are disjunctions of *implicants*, which in turn are conjunctions of *distinct* variables. A *prime* implicant of a formula E is one with a minimal set of variables among all that can appear in DNFs equivalent to E . By absorption, we can retain only the prime implicants. The result is called an *irredundant* DNF (IDNF) of E , and is unique modulo associativity and commutativity. The *co-occurrence graph* of a boolean formula E has its variables as nodes and has an edge between x and y iff they both occur in the same prime implicant of E .

For positive relational queries, the size of the IDNF of the boolean event expressions is polynomial in the size of the table, but often (and necessarily) exponential in the size of the query. This is a good reason for avoiding the explicit computation of the IDNFs, and in particular for not relying on the algorithm in [12]. In recent and independent work Sen et al. [33] proved that for the boolean expressions that arise out of the evaluation of conjunctive queries without self-joins the characterization in [18] can be simplified and one only needs to test whether the co-occurrence graph is a “cograph” [5] which can be done ⁴ in linear time [6].

It is also stated [33] that even for conjunctive queries without self-joins computing co-occurrence graphs likely requires obtaining the IDNF of the boolean expressions. One of our contributions in this paper is to show that an excursion through the IDNF is in fact not necessary because the co-occurrence graphs can be computed directly from the *provenance graph* [25, 14] that captures the computation of the query on a table. Provenance graphs are DAG representations of the event expressions in such a way that most common subexpressions for the entire table (rather than just

⁴Defining cographs seems unnecessary for this paper. It suffices to point out that most cograph recognition algorithms produce (if it exists) something called a “cotree” (sigh) which in the case of co-occurrence graphs associated to boolean formulas is exactly a read-once form!

each tuple) are not replicated. The smaller size of the provenance graphs likely provides practical speedups in the computations (compared for example with the provenance trees of [33]). Moreover, our approach may be applicable to other kinds of queries, as long as their provenance graphs satisfy a simple criterion that we identify.

To give more context to our results, we also note that Hellerstein and Karpinski[21] have shown that if $RP \neq NP$ then deciding whether an arbitrary monotone boolean formula is read-once cannot be done in PTIME in the size of the formula.

The restriction to conjunctive queries without self-joins further allows us to contribute improvements even over an approach that composes our efficient computation of co-occurrence graphs with one of the linear-time algorithms for cograph recognition [6, 19, 3]. Indeed, we show that only a certain subgraph of the co-occurrence graph (we call it the *co-table* graph) is relevant for our stated problem. The co-table graph can be asymptotically smaller than the co-occurrence graph for some classes of queries and instances. To enable the use of only part of the co-occurrence graph we contribute a novel algorithm that computes (when they exist) the read-once forms, using two new ideas: *row decomposition* and *table decomposition*. Using just connectivity tests (eg., DFS), our algorithm is simpler to implement than the cograph recognition algorithms in [6, 19, 3] and it has the potential of affecting the implementation of probabilistic databases.

Moreover, the proof of completeness for our algorithm does not use the cograph characterization on which [33] relies. As such, the algorithm itself provides an alternative new characterization of read-once expressions generated by conjunctive queries without self-joins. This may provide useful insights into extending the approach to handle larger classes of queries.

Having rejected the use of co-occurrence graphs, Sen et al. [33] provide a different approach that derives efficiently the read-once form directly from the computations of the trees underlying the boolean expressions, so called “lineage trees”, by merging read-once forms that correspond to partial formulas. They provide a complexity analysis only for one of the steps that their algorithm applies repeatedly. However, to the best of our understanding of the asymptotic complexity of their algorithm, it appears that our algorithm is faster at least by a multiplicative factor of k^2 where k is the number of tables, and the benefit can often be more.

It is also important to note that neither the results of this paper, nor those of [33] provide complexity dichotomies as does, eg. [8]. It is easy to give a family of probabilistic databases for which the query in Fig. 1(c) generates event expressions of the following form:

$$x_1x_2 + x_2x_3 + \cdots + x_{n-1}x_n + x_nx_{n+1}.$$

These formulas are not read-once, but with a simple memoization (dynamic programming) technique we can compute their probability in time linear in n (see Appendix B).

Roadmap. In Section 2 we review definitions, explain how to compute provenance DAGs for SPJ queries, and compare the sizes of the co-occurrence and co-table graphs. Section 3 presents a characterization of the co-occurrence graphs that correspond to boolean expressions generated by conjunctive queries without self-joins. The characterization uses the provenance DAG. With this characterization we give an efficient algorithm for computing the co-table (and co-occurrence) graph. In Section 4 we give an efficient algorithm that, using the co-table graph, checks if the

result of the query is read-once, and if so computes its read-once form. Putting together these two algorithms we obtain an efficient query-answering algorithm that is *complete* for boolean conjunctive queries without self-joins and tuple-independent databases that yield read-once event expressions. In Section 5 we compare the time complexity of this algorithm with that of Sen et al. [33], and other approaches that take advantage of past work in the read-once and cograph literature. Related work, conclusions and ideas for further work ensue.

2 Preliminaries

A **tuple-independent probabilistic database** is represented by a usual (set-)relational database instance I in which, additionally, every tuple is annotated with a probability in $(0, 1]$, see for example Fig. 1(a). We call this the *probability table representation*. We will denote by $\mathbf{R} = \{R_1, \dots, R_k\}$ the relational schema of the representation. By including/excluding each tuple independently with probability of its annotation, the representation defines a set of \mathbf{R} -instances called *possible worlds* and the obvious probability distribution on this set, hence a discrete probability space. For a given tuple $t \in R_i$ this space's event “ t occurs” (the set of possible worlds in which t occurs) has probability exactly the annotation of t in the representation. Following the intensional approach [27, 11, 36] we also consider the *event table representation* which consists of the same tables, but in which every tuple is annotated by its unique tuple id, for example see Fig. 1(b).

The **tuple ids** play three distinct but related roles: (1) they identify tuples uniquely over *all* tables and in fact we will often call the tuple ids just tuples, (2) they are boolean variables, (3) they denote the events “the tuple occurs” in the probability space of all possible worlds. The last two perspectives can be combined by saying that the tuple ids are boolean-valued random variables over said probability space. Moreover, an *event expression* is a boolean expression with the tuple ids as variables.

The intensional approach further defines the semantics of the **relational algebra operators** on event tables, i.e., relational instances in which the tuples are annotated with event expressions. In this paper we will only need monotone boolean expressions because our queries only use selection, projection and join and these operators do not introduce negation. Otherwise, joins produce conjunctions, projections produce disjunctions, and selections erase the non-compliant tuples. It is worth observing that the relational algebra on event tables is essentially a particular case of the algebra on c-tables [22], and precisely a particular case of the relational algebra on semiring-annotated relations [15].

Since by now they are well understood (see the many papers we cited so far), we do not repeat here the definition of select, project, and join on tables annotated with boolean event expressions but instead we explain how they produce **provenance graphs**. The concept that we define here is a small variation on the provenance graphs defined in [25, 14] where conjunctive queries (part of mapping specifications) are treated as a black box. It is important for the provenance graphs used in this paper to reflect the structure of different SPJ query plans that compute the same conjunctive query.

A provenance graph (PG) is a directed acyclic graph (DAG) H such that the nodes $V(H)$ of H are labeled by variables or by the operation symbols \cdot and $+$. As we show below, each node

corresponds to a tuple in an event table that represents the set of possible worlds of either the input database or some intermediate database computed by the query plan. An edge $u \rightarrow v$ is in $E(H)$ if the tuple corresponding to u is computed using the tuple corresponding to v in either a join (in which case u is labeled with \cdot) or a projection (in which case u is labeled with $+$). The nodes with no outgoing edges are those labeled with variables and are called leaves while the nodes with no incoming edges are called roots (and can be labeled with either operation symbol). Provenance graphs (PGs) are closely related to the *lineage trees* of [33]. In fact, the lineage trees are tree representations of the boolean event expressions, while PGs are more economical: they represent the same expressions but without the multiplicity of common subexpressions. Thus, they are associated with an entire table rather than with each tuple separately, each root of the graph corresponding to a tuple in the table.⁵

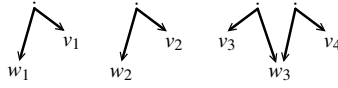


Figure 2: Provenance graph for $R \bowtie S$.

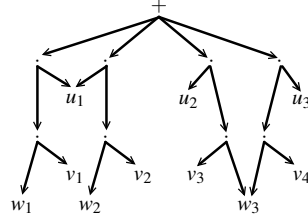


Figure 3: Provenance graph for $\pi_0((R \bowtie S) \bowtie T)$.

We explain how the SPJ algebra works on tables with PGs. If tables R_1 and R_2 have PGs H_1 and H_2 then the PG for $R_1 \bowtie R_2$ is constructed as follows. Take the disjoint union H of H_1 and H_2 . For every $t_1 \in R_1$ and $t_2 \in R_2$ that do join, add a new root labeled with \cdot and make the root of H_1 corresponding to t_1 and that of H_2 corresponding to t_2 children of this new root. Afterwards, delete (recursively) any remaining roots from H_1 and H_2 . For example, referring again to Fig. 1, the PG associated with the table computed by $R \bowtie S$ is shown in Fig. 2.

For selection, delete (recursively) the roots that correspond to the tuples that do not satisfy the selection predicate. For projection, consider a table T with PG H and X a subset of its attributes. The PG for $\pi_X R$ is constructed as follows. For each $t \in \pi_X R$, let t_1, \dots, t_m be all the tuples in R that X -project to t . Add to H a new root labeled with $+$ and make the roots in H corresponding to t_1, \dots, t_m the children of this new root. Referring again to Fig. 1, the PG associated with the result of the query plan $\pi_0((R \bowtie S) \bowtie T)$ is shown in Fig. 3. Since the query is boolean, this PG has just one root.

⁵Note that to facilitate the comparison with the lineage trees the edge direction here is the opposite of the direction in [25, 14].

The boolean event expressions that annotate tuples in the event tables built in the intensional approach can be read off the provenance graphs. Indeed, if t occurs in an (initial, intermediate, or final) table T whose PG is H , then, starting at the root u of H corresponding to t , traverse the subgraph induced by all the nodes reachable from u and build the boolean expression recursively using parent labels as operation symbols and the subexpressions corresponding to the children as operands. For example, we read $((w_1 \cdot v_1) \cdot u_1) + (u_1 \cdot (w_2 \cdot v_2)) + (u_2 \cdot (v_3 \cdot w_3)) + ((w_3 \cdot v_4) \cdot u_3)$ off the PG in Fig. 3.

The focus of this paper is the case when the boolean (event) expressions are **read-once**, i.e., they are equivalent to expressions in which every variable occurs exactly once, the latter said to be in *read-once form*. For boolean expressions that are read-once, the read-once form is unique (modulo associativity and commutativity)⁶. The interest in read-once formulas derives from the fact that in tuple-independent databases the tuples in the input representation occur *independently* in possible worlds. More complex boolean expressions denote events whose probability needs to be computed from the probabilities of the variables, i.e., the probabilities of the independent “tuple occurs” events in the input. When such event expressions are in read-once form their probability can be computed efficiently in linear time in the number of variables using the rules (3) and (4) in Section 1.

Given a tuple-independent probabilistic database and an SPJ query plan, hence the resulting provenance graph, our objective in this paper is to decide efficiently when the boolean expression(s) read off the PG are read-once, and when they are, to compute their read-once form(s) efficiently, hence the associated probability(es).

In this paper we consider only **boolean conjunctive queries**. We can do this without loss of generality because we can associate to a non-boolean conjunctive query Q and an instance I a set of boolean queries in the usual manner: for each tuple t in the answer relation $Q(I)$, consider the boolean conjunctive query Q_t which is obtained from Q by replacing the head variables with the corresponding values in t . Note that the PGs that result from boolean queries have *exactly one root*. We will also use $Q(I)$ to denote the boolean expression generated by evaluating the query Q on instance I , which may have different (but equivalent) forms based on the query plan.

Moreover, we consider only queries **without self-join**. Therefore our queries have the form

$$Q() : -R_1(\mathbf{x}_1), \dots, R_k(\mathbf{x}_k)$$

where R_1, \dots, R_k are all *distinct* table names while the \mathbf{x}_i ’s are tuples of FO variables⁷ or constants, possibly with repetitions, matching the arities of the tables. If the database has tables that do not appear in the query, they are of no interest, so we will always assume that our queries feature all the table names in the database schema \mathbf{R} .

As we have stated above, we only need to work with *monotone* boolean formulas (all literals are positive, only disjunction and conjunction operations). Every such formula is equivalent to (many) **disjunctive normal forms** (DNFs) which are disjunctions of conjunctions of variables.

⁶This seems to have been known for a long time. We could not find an explicitly stated theorem to this effect in the literature, but, for example, it is clear that the result of the algorithm in [12] is uniquely determined by the input.

⁷FO (first-order) is to emphasize the distinction between the variables in the query subgoals and the variables in the boolean expressions.

These conjunctions are called *implicants* for the DNF. By idempotence we can take the variables in an implicant to be distinct and the implicants of a DNF to be distinct from each other. A *prime* implicant of a formula f is one with a minimal set of variables among all that can appear in DNFs equivalent to f . By absorption, we can retain only the prime implicants in a DNF. The result is called *the irredundant DNF* (IDNF) of f , as it is uniquely determined by f (modulo associativity and commutativity). We usually denote it by f_{IDNF} . Note that in particular the set of prime implicants is uniquely determined by f .

The **co-occurrence graph**, notation G_{co} , of a boolean formula f is an undirected graph whose set of vertices $V(G_{co})$ is the set $\text{Var}(f)$ of variables of f and whose set $E(G_{co})$ of edges is defined as follows: there is an edge between x and y iff they both occur in the same prime implicant of f . Therefore, G_{co} is uniquely determined by f and it can be constructed from f_{IDNF} . This construction is quadratic in the size of f_{IDNF} but of course f_{IDNF} can be exponentially larger than f . Fig. 4 shows the co-occurrence graph for the boolean expression f in equation (1) of Example 1. As this figure shows, the co-occurrence graphs for expressions generated by conjunctive queries without self join are always k -partite⁸ graphs on tuple variables from k different tables.

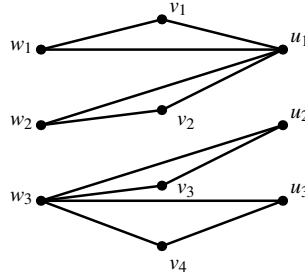


Figure 4: G_{co} for f in Example 1.

We are interested in the co-occurrence graph G_{co} of a boolean formula f because it plays a crucial role in f being read-once. Indeed [18] has shown that a monotone f is read-once iff (1) it is “normal” and (2) its G_{co} is a “cograph”. We don’t need to discuss normality because [33] has shown that for formulas that arise from conjunctive queries without self-joins it follows from the cograph property. We will also avoid defining what a cograph is (see [5, 6]) except to note that cograph recognition can be done in linear time [6, 19, 3] and that when applied to the co-occurrence graph of f the recognition algorithms also produce, in effect, the read-once form of f , when it exists.

Although the co-occurrence graph of f is defined in terms of f_{IDNF} , we show in Section 3 that when f is the event expression produced by a boolean conjunctive query without self-joins then we can efficiently compute the G_{co} of f from the provenance graph H of any plan for the query. Combining this with any of the cograph recognition algorithms we just cited, this yields one algorithm for the goal of our paper, which we will call a **cograph-help algorithm**.

Because it uses the more general-purpose step of cograph recognition a cograph-help algorithm will not fully take advantage of the restriction to conjunctive queries without self-joins. Intuitively,

⁸A graph $(V_1 \cup \dots \cup V_k, E)$ is k -partite, if for any edge $(u, v) \in E$ where $u \in V_i$ and $v \in V_j$, $i \neq j$.

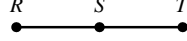


Figure 5: G_T for the relations in Example 1.

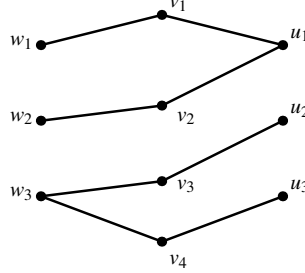


Figure 6: G_C for f in Example 1.

with this restriction there may be lots of edges in G_{co} that are irrelevant because they link tuples that are not joined by the query. This leads us to the notion of co-table graph defined below.

Toward the definition of the co-table graph we also need that of **table-adjacency graph**, notation G_T . Given a boolean query without self-joins $Q() : -R_1(\mathbf{x}_1), \dots, R_k(\mathbf{x}_k)$ the vertex set $V(G_T)$ is the set of k table names R_1, \dots, R_k . We will say that R_i and R_j are *adjacent* iff \mathbf{x}_i and \mathbf{x}_j have at least one FO variable in common i.e., R_i and R_j are joined by the query. The set of edges $E(G_T)$ consists of the pairs of adjacent table names. The table-adjacency graph G_T for the query in Example 1 is depicted in Fig. 5.

The table-adjacency graph G_T helps us remove edges irrelevant to a query from the graph G_{co} . For example, if there is an edge between $x \in R_i$ and $x' \in R_j$ in G_{co} , but there is no edge between R_i and R_j in G_T , then (i) either there is no path connecting R_i to R_j in G_T (so all tuples in R_i pair with all tuples in R_j), or, (ii) x and x' are connected in G_{co} via a set of tuples x_1, \dots, x_ℓ , such that the tables containing these tuples are connected by a path in G_T . Our algorithm in Section 4 shows that all such edges (x, x') can be safely deleted from G_{co} for the evaluation of the query that yielded G_T .

Definition 1. The **co-table graph** G_C is the subgraph of G_{co} with $V(G_C) = V(G_{co})$ and such that given two tuples $x \in R_i$ and $x' \in R_j$ there is an edge $(x, x') \in E(G_C)$ iff $(x, x') \in E(G_{co})$ and R_i and R_j are adjacent in G_T .

The co-table graph G_C generated by the event tables and query in Fig. 1 is shown in Fig. 6 (it is *not* a cograph!).

Co-occurrence graph vs. co-table graph The advantage of using the co-table graph instead of the co-occurrence graph is most dramatic in the following example:

Example 3. Consider $Q() : -R_1(\mathbf{x}_1), R_2(\mathbf{x}_2)$ where \mathbf{x}_1 and \mathbf{x}_2 have no common FO variable. Assuming that each of the tables R_1 and R_2 has n tuples, G_{co} has n^2 edges while G_C has none. A cograph-help algorithm must spend $\Omega(n^2)$ time even if it only reads G_{co} !

On the other hand, G_C can be as big as G_{co} . In fact, when G_T is a complete graph (see next example), $G_C = G_{co}$.

Example 4. Consider $Q() : -R_1(\mathbf{x}_1, y), \dots, R_k(\mathbf{x}_k, y)$ where \mathbf{x}_i and \mathbf{x}_j have no common FO variable if $i \neq j$. Here G_T is the complete graph on R_1, \dots, R_k and $G_C = G_{co}$.

However, it can be verified that both our algorithm and the cograph-help algorithm have the same time complexity on the above example.

3 Computing the Co-Table Graph

In this section we show that given as input the provenance DAG H of a boolean conjunctive query plan without self-joins Q on a table-independent database representation I , the co-table graph G_C and the co-occurrence graph G_{co} of the boolean formula $Q(I)$ (see definitions in section 2) can be computed in poly-time in the sizes of H, I and Q .

It turns out that G_C and G_{co} are computed by similar algorithms, one being a minor modification of the other. As discussed in section 1, the co-occurrence graph G_{co} can then be used in conjunction with cograph recognition algorithms (eg., [6, 19, 3]), to find the read-once form of $Q(I)$ if it exists. On the other hand, the smaller co-table graph G_C is used by our algorithm described in section 4 for the same purpose.

We use $\text{Var}(f)$ to denote the sets of variables in a monotone boolean expression f . Recall that the provenance DAG H is a layered graph where every layer corresponds to a select, project or join operation in the query plan. We define the *width* of H as the maximum number of nodes at any layer of the DAG H and denote it by β_H . The main result in this section is summarized by the following theorem.

Theorem 1. Let $f = Q(I)$ be the boolean expression computed by the query plan Q on the table representation I (f can also be read off the provenance graph of Q on I , H), $n = |\text{Var}(f)|$ be the number of variables in f , $m_H = |E(H)|$ be the number of edges of H , β_H be the width of H , and $m_{co} = |E(G_{co})|$ be the number of edges of G_{co} , the co-occurrence graph of f .

1. G_{co} can be computed in time $O(nm_H + \beta_H m_{co})$.
2. Further, the co-table graph G_C of f can be computed in time $O(nm_H + \beta_H m_{co} + k^2 \alpha \log \alpha)$ where k is the number of tables in Q , and α is the maximum arity (width) of the tables in Q .

3.1 LCA-Based Characterization of the Co- Occurrence Graph

Here we give a characterization of the presence of an edge (x, y) in G_{co} based on the *least common ancestors* of x and y in the graph H .

Again, let $f = Q(I)$ be the boolean expression computed by the query plan Q on the table representation I . As explained in section 2 f can also be read off the provenance graph H of Q and I since H is the representation of f without duplication of common subexpressions.

The absence of self-joins in Q implies the following.

Lemma 1. *The DNF generated by expanding f (or H) using only the distributivity rule is in fact the IDNF of f up to idempotency (i.e. repetition of the same prime implicant is allowed).*

Proof. Let g be the DNF generated from f by applying distributivity repeatedly. Due to the absence of self-joins g every implicant in g will have exactly one tuple from every table. Therefore, for any two implicants in g the set of variables in one is not a strict subset of the set of variables in the other and further absorption (eg., $xy + xyz = xy$) does not apply. (At worst, two implicants can be the same and the idempotence rule reduces one.) Therefore, g is also irredundant and hence *the* IDNF of f (up to commutativity and associativity). \square

Denote by f_{IDNF} the IDNF of f , which, as we have seen, can be computed from f just by applying distributivity.

As with any DAG, we can talk about the nodes of H in terms of successors, predecessors, ancestors, and descendants, and finally about the *least common ancestors* of two nodes, denoted $lca(x, y)$. Because H has a root $lca(x, y)$ is never empty. When H is a tree, $lca(x, y)$ consists of a single node. For a node $u \in V(H)$, we denote the set of leaf variables which are descendants of u by $Var(u)$ (overloaded notation warning!); in other words, a variable x belongs to $Var(u)$, $u \in V(H)$, if and only if x is reachable from u in H . Now we prove the key lemma of this section:

Lemma 2. *Two variables $x, y \in Var(f)$ belong together to a (prime) implicant of f_{IDNF} if and only if the set $lca(x, y)$ contains a \cdot -node.*

Proof. (if) Suppose $lca(x, y)$ contains a \cdot -node u , i.e., x, y are both descendants of two distinct successors v_1, v_2 of u . Since the \cdot operation multiplies all variables in $Var(v_1)$ with all variables in $Var(v_2)$, x and y will appear together in some implicant in f_{IDNF} which will not be absorbed by other implicants by Lemma 1.

(only if) Suppose that x, y appear together in an implicant of f_{IDNF} and $lca(x, y)$ contains no \cdot -node. Then no \cdot -node in $V(H)$ has x, y in $Var(v_1), Var(v_2)$, where v_1, v_2 are its two distinct successors (note that any \cdot -node in a provenance DAG H can have exactly two successors). This implies that x and y can never be multiplied, contradiction. \square

Since there are exactly k tables in the query plan, every implicant in f_{IDNF} will be of size k . Therefore:

Lemma 3. *For every variable $x \in Var(f)$ and \cdot -node $u \in V(H)$, if $x \in Var(u)$, then $x \in Var(v)$ for exactly one successor v of u .*

Proof. If $x \in Var(f)$ belongs to $Var(v_1), Var(v_2)$ for two distinct successors v_1, v_2 of u , then some implicant in f_{IDNF} will have $< k$ variables since $x \cdot x = x$ by idempotence. \square

The statement of Lemma 2 provides a criterion for computing G_{co} using the computation of least common ancestors in the provenance graph, which is in often more efficient than computing the entire IDNF. We have shown that this criterion is satisfied in the case of conjunctive queries without self-joins. But it may also be satisfied by other kinds of queries, which opens a path to identifying other cases in which such an approach would work.

3.2 Computing the Table-Adjacency Graph

It is easier to describe the computation of G_T if we use the query in rule form $Q() : -R_1(\mathbf{x}_1), \dots, R_k(\mathbf{x}_k)$.

The rule form can be computed in linear time from the SPJ query plan. Now the vertex set $V(G_T)$ is the set of table names R_1, \dots, R_k . and an edge exists between R_i, R_j iff \mathbf{x}_i and \mathbf{x}_j have at least one FO variable in common i.e., R_i and R_j are joined. Whether or not such an edge should be added can be decided in time $O(\alpha \log \alpha)$ by sorting and intersecting \mathbf{x}_i and \mathbf{x}_j . Here α is the maximum arity (width) of the tables R_1, \dots, R_k . Hence G_T can be computed in time $O(k^2 \alpha \log \alpha)$.

3.3 Computing the Co-Table Graph

Recall that co-table graph G_C is a subgraph of the co-occurrence graph G_{co} where we add an edge between two variables x, y , only if the tables containing these two tuples are adjacent in the table-adjacency graph G_T . Algorithm 1 COMPCOTABLE constructs the co-table graph G_C by a single *bottom-up* pass over the graph H .

Algorithm 1 *Algorithm COMPCOTABLE*

Input: Query plan DAG H and table-adjacency graph G_T

Output: Co-table graph G_C .

```

1: – Initialize  $V(G_C) = \text{Var}(f)$ ,  $E(G_C) = \phi$ .
2: – For all variables  $x \in \text{Var}(f)$ , set  $\text{Var}(x) = \{x\}$ .
3: – Do a topological sort on  $H$  and reverse the sorted order.
4: for every node  $u \in V(H)$  in this order do
5:   /* Update  $\text{Var}(u)$  set for both +-node and --node  $u$  */
6:   – Set  $\text{Var}(u) = \bigcup_v \text{Var}(v)$ , where the union is over all successors  $v$  of  $u$ .
7:   if  $u \in V(H)$  is a --node then
8:     /* Add edges to  $G_C$  only for a --node */
9:     – Let  $v_1, v_2$  be its two successors.
10:    for every two variables  $x \in \text{Var}(v_1)$  and  $y \in \text{Var}(v_2)$  do
11:      if (i) the tables containing  $x, y$  are adjacent in  $G_T$  and (ii) the edge  $(x, y)$  does not exist
        in  $E(G_C)$  yet then
12:        – Add an edge between  $x$  and  $y$  in  $E(G_C)$ .
13:      end if
14:    end for
15:  end if
16: end for

```

It is easy to see that a minor modification of the same algorithm can be used to compute the co-occurrence graph G_{co} : in Step 11 we simply skip the check whether the tables containing the two tuples are adjacent in G_T . Since this is the only place where G_T is used, the time for the computation of G_C does not include the time related to computing/checking G_T .

Correctness. By a simple induction, it can be shown that the set $\text{Var}(u)$ is correctly computed at every step, i.e., it contains the set of all nodes which are reachable from u in H (since the nodes are processed in reverse topological order and $\text{Var}(u)$ is union of $\text{Var}(v)$ for over all successors v of u). Next lemma shows that algorithm COMPCoTABLE correctly builds the co-table graph G_C (proof in Appendix C).

Lemma 4. *Algorithm COMPCoTABLE adds an edge (x, y) to G_C if and only if x, y together appear in some implicant in f_{IDNF} and the tables containing x, y are adjacent in G_T .*

Time Complexity. Here we give a sketch of the time complexity analysis, details can be found in the appendix (Section C.1). Computation of the table adjacency graph takes $O(k^2 \alpha \log \alpha)$ time as shown in Section 3.2. The total time complexity of algorithm COMPCoTABLE as given in Theorem 1 is mainly due to two operations: (i) computation of the $\text{Var}(u)$ set at every internal node $u \in V(H)$, and (ii) to perform the test for pairs x, y at two distinct children of a \cdot -node, whether the edge (x, y) already exists in G_C , and if not, to add the edge.

We show that the total time needed for the first operation is $O(nm_H)$ in total: for every internal node $u \in V(H)$ we can scan the variables sets of all its immediate successor in $O(nd_u)$ time to compute $\text{Var}(u)$, where d_u is the outdegree of node u in H . This gives total $O(nm_H)$ time. On the other hand, for adding edges (x, y) in G_C , it takes total $O(m_{co}\beta_H)$ time during the execution of the algorithm, where m_{co} is the number of edges in the co-occurrence graph (and not in the co-table graph, even if we compute the co-table graph G_C) and β_H is the width of the graph H . To show this, we show that two variables x, y are considered by the algorithm at Step 10 if and only if the edge (x, y) already exists in the co-occurrence graph G_{co} , however, the edge may not be added to the co-table graph G_C if the corresponding tables are not adjacent in the table adjacency graph G_T . We also show that any such edge (x, y) will be considered at a unique level of the DAG H . In addition to these operations, the algorithm does initialization and a topological sort on the vertices which take $O(m_H + n_H)$ time ($n_H = |V(H)|$) and are dominated by the these two operations.

4 Computing the Read-Once Form

Our algorithm COMPRO (for *Compute Read-Once*) takes an instance I of the schema $\mathbf{R} = R_1, \dots, R_k$, a query $Q() : -R_1(\mathbf{x}_1), R_2(\mathbf{x}_2), \dots, R_k(\mathbf{x}_k)$ along with the table adjacency graph G_T and co-table graph G_C computed in the previous section as input, and outputs whether $Q(I)$ is read-once. (if so it computes its unique read-once form).

Theorem 2. *Suppose we are given a query Q , a table-independent database representation I , the co-table graph G_C and the table-adjacency graph G_T for Q on I as inputs. Then*

1. *Algorithm COMPRO decides correctly whether the expression generated by evaluating Q on I is read-once, and if yes, it returns the unique read-once form of the expression, and,*
2. *Algorithm COMPRO runs in time $O(m_T \alpha \log \alpha + (m_C + n) \min(k, \sqrt{n}))$,*

where $m_T = |E(G_T)|$ is the number of edges in G_T , $m_C = |E(G_C)|$ is the number of edges in G_C , n is the total number of tuples in I , k is the number of tables, and α is the maximum size of any subgoal.

4.1 Algorithm COMPRO

In addition to the probabilistic database with tables R_1, \dots, R_k and input query Q , our algorithm also takes the *table-adjacency graph* G_T and the *co-table graph* G_C computed in the first phase as discussed in Section 3. The co-table graph G_C also helps us to remove *unused tuples* from all the tables which do not appear in the final expression – every unused tuple won’t have a corresponding node in G_C . So from now on we can assume wlog. that every tuple in every table appears in the final expression f .

The algorithm COMPRO uses two decomposition operations: *Row decomposition* is a *horizontal* decomposition operation which partitions the rows or tuples in every table into the same number of groups and forms a set of sub-tables from every table. On the other hand, *Table decomposition* is a *vertical* decomposition operation. It partitions the set of tables into groups and a *modified sub-query* is evaluated in every group. For convenience, we will represent the instance I as $R_1[T_1], \dots, R_k[T_k]$, where T_i is the set of tuples in table R_i . Similarly, for a subset of tuples $T'_i \subseteq T_i$, $R_i[T'_i]$ will denote the instance of relation R_i containing exactly the tuples in T'_i . The algorithm COMPRO is given in Algorithm 2.

Row Decomposition. The row decomposition operation partitions the tuples variables in *every* table into ℓ disjoint groups. In addition, it decomposes the co-table graph G_C into $\ell \geq 2$ disjoint *induced subgraphs*⁹ corresponding to the above groups. For every pair of distinct groups j, j' , and for every pair of distinct tables $R_i, R_{i'}$, no tuple in group j of R_i ever joins with a tuple in group j' of $R_{i'}$ (recall that the query does not have any self-join operation). The procedure for row decomposition is given in Algorithm 3¹⁰.

Table Decomposition. On the other hand, the table decomposition operation partitions the set of tables $\mathbf{R} = R_1, \dots, R_k$ into $\ell \geq 2$ disjoint groups $\mathbf{R}_1, \dots, \mathbf{R}_\ell$. It also decomposes the table-adjacency graph G_T and co-table graph G_C into ℓ disjoint induced subgraphs $G_{T,1}, \dots, G_{T,\ell}$, and, $G_{C,1}, \dots, G_{C,\ell}$ respectively corresponding to the above groups. The groups are selected in such a way that all tuples in the tables in one group join with all tuples in the tables in another group. This procedure also modifies the sub-query to be evaluated on every group by making the subqueries of different groups mutually independent by introducing free variables, i.e., they do not share any common variables after a successful table decomposition. Algorithm 4 describes the table decomposition operation. Since the table decomposition procedure changes the input query Q to $\widehat{Q} = \widehat{Q}_1, \dots, \widehat{Q}_\ell$, it is crucial to ensure that changing the query to be evaluated does not change the answer to the final expression. This is shown in Lemma 11 in Appendix D.

⁹A subgraph H of G is an induced subgraph, if for any two vertices $u, v \in V(H)$, if $(u, v) \in E(G)$, then $(u, v) \in E(H)$.

¹⁰It should be noted that the row decomposition procedure may be called on $R_{i_1}[T'_{i_1}], \dots, R_{i_p}[T'_{i_p}]$ and G'_C , where R_{i_1}, \dots, R_{i_p} is a subset of the relations from R_1, \dots, R_k , $T'_{i_1}, \dots, T'_{i_p}$ are subsets of the respective set of tuples T_{i_1}, \dots, T_{i_p} , and G'_C is the induced subgraph of G_C on $T'_{i_1}, \dots, T'_{i_p}$. For simplicity in notations, we use $R_1[T_1], \dots, R_k[T_k]$. This holds for table decomposition as well.

Algorithm 2 $\text{COMPRO}(Q, I = \langle R_1[T_1], \dots, R_k[T_k] \rangle, G_C, G_T, \text{FLAG})$

Input: Query Q , tables $R_1[T_1], \dots, R_k[T_k]$, co-table graph G_C , table-adjacency graph G_T , and a boolean parameter **FLAG** which is true if and only if row decomposition is performed at the current step.

Output: If successful, the unique read-once form f^* of the expression for $Q(I)$

```

1: if  $k = 1$  then
2:   return  $\sum_{x \in T_1} x$  with success. (/* all unused tuples are already removed */)
3: end if
4: if  $\text{FLAG} = \text{TRUE}$  then /* Row decomposition */
5:   – Perform  $\text{RD}(\langle R_1[T_1], \dots, R_k[T_k] \rangle, G_C)$ .
6:   if row decomposition returns with success then /* RD partitions every table and  $G_C$  into  $\ell \geq 2$  disjoint groups */
7:     – Let the groups returned be  $\langle \langle T_1^j, \dots, T_k^j \rangle, G_{C,j} \rangle, j \in [1, \ell]$ .
8:     –  $\forall j \in [1, \ell]$ , let  $f_j = \text{COMPRO}(Q, \langle R_1[T_1^j], \dots, R_k[T_k^j] \rangle, G_{C,j}, G_T, \text{FALSE})$ .
9:     return  $f^* = f_1 + \dots + f_\ell$  with success.
10:  end if
11: else /* Table decomposition */
12:   – Perform  $\text{TD}(\langle R_1[T_1], \dots, R_k[T_k] \rangle, Q, G_T, G_C)$ .
13:   if table decomposition returns with success then /* TD partitions  $I, G_C$  and  $G_T$  into  $\ell \geq 2$  disjoint groups,  $\sum_{j=1}^\ell k_j = k$  */
14:     Let the groups returned be  $\langle \langle R_{j,1}, \dots, R_{j,k_j} \rangle, \widehat{Q}_j, G_{C,j}, G_{T,j} \rangle, j \in [1, \ell]$ .
15:     –  $\forall j \in [1, \ell]$ ,  $f_j = \text{COMPRO}(\widehat{Q}_j, \langle R_1[T_1], \dots, R_k[T_k] \rangle, G_{C,j}, G_{T,j}, \text{TRUE})$ .
16:     return  $f^* = f_1 \cdot \dots \cdot f_\ell$  with success.
17:   end if
18: end if
19: if the current operation is not successful then /* Current row or table decomposition is not successful and  $k > 1$  */
20:   return with failure: “ $Q(I)$  is not read-once”.
21: end if

```

The following lemma shows that if row-decomposition is successful, then table decomposition cannot be successful and vice versa. However, both of them may be unsuccessful in case the final expression is not read-once. The proof of the lemma is in Appendix D).

Lemma 5. *At any step of the recursion, if row decomposition is successful then table decomposition is unsuccessful and vice versa.*

Therefore, in the top-most level of the recursive procedure, we can verify which operation can be performed – if both of them fail, then the final expression is not read-once which follows from the correctness of our algorithm. If the top-most recursive call performs a successful row decomposition initially the algorithm COMPRO is called as $\text{COMPRO}(Q, \langle R_1[T_1], \dots, R_k[T_k] \rangle, G_C, G_T, \text{TRUE})$. The last boolean argument is **TRUE** if and only if row decomposition is performed at

Algorithm 3 $\text{RD}(\langle R_1[T_1], \dots, R_k[T_k] \rangle, G'_C)$

Input: Tables $R_1[T_1], \dots, R_k[T_k]$, and induced subgraph G'_C of G_C on $\bigcup_{i=1}^k T_i$

Output: If successful, the partition of G'_C and tuple variables of every input tables into $\ell \geq 2$ connected components: $\langle \langle T_{1,j}, \dots, T_{k,j} \rangle, G'_{C,j} \rangle, j \in [1, \ell]$

- 1: – Run BFS or DFS to find the connected components in G'_C .
 - 2: – Let ℓ be the number of connected components.
 - 3: **if** $\ell = 1$ **then** $\{/* \text{there is only one connected component} */\}$
 - 4: **return** with failure: “Row decomposition is not possible”.
 - 5: **else**
 - 6: – Let the tuples (vertices) of table R_i in the j -th connected component j of G'_C be $T_{i,j}$
 - 7: – Let the induced subgraph for connected component j be $G_{C,j}$.
 - 8: **return** $\langle \langle T_{1,1}, \dots, T_{k,1} \rangle, G'_{C,1} \rangle, \dots, \langle \langle T_{1,\ell}, \dots, T_{k,\ell} \rangle, G'_{C,\ell} \rangle$ with success.
 - 9: **end if**
-

the current level of the recursion tree. If in the first step table decomposition is successful, then the value of the last boolean variable in the initial call will be FALSE.

Correctness. The following two lemmas respectively show the soundness and completeness of the algorithm COMPRO (proofs are in Appendix D).

Lemma 6. (Soundness) *If the algorithm returns with success, then the expression f^* returned by the algorithm COMPRO is equivalent to the expression $Q(I)$ generated by evaluation of query Q on instance I . Further, the output expression f^* is in read-once form.*

Lemma 7. (Completeness) *If the expression $Q(I)$ is read-once, then the algorithm COMPRO returns the unique read-once form f^* of the expression.*

For completeness, it suffices to show that if $Q(I)$ is read-once, then the algorithm does not exit with error. Indeed, if the algorithm returns with success, as showed in the soundness lemma, the algorithm returns an expression f^* in read-once form which is the unique read-once form of $Q(I)$ [12],[5].

Time Complexity. Consider the recursion tree of the algorithm COMPRO. Lemma 5 shows that at any level of the recursion tree, either all recursive calls use the row decomposition procedure, or all recursive calls use the column decomposition procedure. The time complexity of COMPRO given in Theorem 2 is analyzed in the following steps. If n' = the total number of input tuples at the current recursive call and m'_C = the number of edges in the induced subgraph of G'_C on these n' vertices, we show that row decomposition takes $O(m'_C + n')$ time and, *not considering the time needed to compute the modified queries \widehat{Q}_j* (Step 13 in Algorithm 4), the table decompositions procedure takes $O(m'_C + n')$ time. Then we consider the time needed to compute the modified queries and show that these steps over *all* recursive calls of the algorithm take $O(m_T \alpha \log \alpha)$ time in total, where α is the maximum size of a subgoal in the query Q . Finally, we give a bound of $O(\min(k, \sqrt{n}))$ on the height of the recursive tree for the algorithm COMPRO. However, note that

Algorithm 4 TD($\langle R_1[T_1], \dots, R_k[T_k] \rangle, \langle Q() : -R_1(\mathbf{x}_1), \dots, R_k(\mathbf{x}_k) \rangle, G'_C, G'_T$)

Input: Tables $R_1[T_1], \dots, R_k[T_k]$ **query** $Q() : -R_1(\mathbf{x}_1), \dots, R_k(\mathbf{x}_k)$ **induced subgraph** G'_T of G_T **on** $\bigcup_{i=1}^k R_i$, **induced subgraph** G'_C of G_C **on** $\bigcup_{i=1}^k T_i$

Output: If successful, a partition of input tables, G'_T, G'_C into ℓ groups, and an updated subquery for every group

```

1: for all edges  $e = (R_i, R_j)$  in  $G'_T$  do
2:   – Annotate the edge  $e$  with common variables  $C_e$  in the vectors  $\mathbf{x}_i, \mathbf{x}_j$ .
3:   – Mark the edge  $e$  with a “+” if for every pair of tuple variables  $x \in T_i$  and  $y \in T_j$ , the edge  $(x, y)$  exists in  $G'_C$ . Otherwise mark the edge with a “–”.
4: end for
5: – Run BFS or DFS to find the connected components in  $G_T$  w.r.t “–” edges
6: – Let  $\ell$  be the number of connected components.
7: if  $\ell = 1$  then  $\{ /* there is only one connected component */ \}$ 
8:   return with “Failure: Table decomposition is not possible”.
9: else
10:  – Let  $G'_{T,1}, \dots, G'_{T,\ell}$  be the induced subgraphs of  $\ell$  connected components of  $G'_T$  and  $G'_{C,1}, \dots, G'_{C,\ell}$  be the corresponding induced subgraph for  $G'_C$ .
11:  – Let  $\mathbf{R}_p = \langle R_{p,1}, \dots, R_{p,k_p} \rangle$  be the subset of tables in the  $p$ -th component of  $G'_T$ ,  $p \in [1, \ell]$ .
12:   $/* Compute a new query for every component */$ 
13:  for every component  $p$  do
14:    for every table  $R_i$  in this component  $p$  do
15:      – Let  $C_i = \bigcup_e C_e$  be the union of common variables  $C_e$  over all edges  $e$  from  $R_i$  to tables in different components of  $G'_T$  (all such edges are marked with ‘+’)
16:      – For every common variable  $z \in C_i$ , generate a new (free) variable  $z^i$ , and replace all occurrences of  $z$  in vector  $\mathbf{x}_i$  by  $z^i$ . Let  $\widehat{\mathbf{x}}_i$  be the new vector.
17:      – Change the query subgoal for  $R_i$  from  $R_i(\mathbf{x}_i)$  to  $R_i(\widehat{\mathbf{x}}_i)$ .
18:    end for
19:    Let  $\widehat{Q}_p() : -R_{p,1}(\widehat{\mathbf{x}}_{p,1}), \dots, R_{p,k_p}(\widehat{\mathbf{x}}_{p,k_p})$  be the new query for component  $p$ .
20:  end for
21:  return  $\langle \langle R_{1,1}, \dots, R_{1,k_1} \rangle, \widehat{Q}_1, G_{C,1}, G_{T,1} \rangle, \dots, \langle \langle R_{\ell,1}, \dots, R_{\ell,k_\ell} \rangle, \widehat{Q}_\ell, G_{C,\ell}, G_{T,\ell} \rangle$  with success.
22: end if

```

at every step, for row or table decomposition, every tuple in G'_C goes to exactly one of the recursive calls, and every edge in G'_C goes to at most one of the recursive calls. So for both row and table decomposition at every level of the recursion tree the total time is $O(m_C + n)$. Combining all these observations, the total time complexity of the algorithm is $O(m_T \alpha \log \alpha + (m_C + n) \min(k, \sqrt{n}))$ as stated in Theorem 2. The details can be found in Appendix D.1).

Example. Here we illustrate our algorithm. Consider the query Q and instance I from Example 1 in the introduction. The input query is $Q() : -R(x)S(x,y)T(y)$. In the first phase, the table-adjacency graph G_T and the co-table graph G_C are computed. These graphs are depicted in

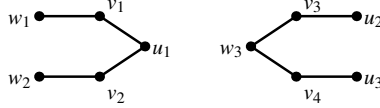


Figure 7: $G_{C,1}$ and $G_{C,2}$.

Fig. 5 and Fig. 6 respectively.

Now we apply COMPRO. There is a successful row decomposition at the top-most recursive call that decomposes G_C into the two subgraphs $G_{C,1}, G_{C,2}$ shown in Fig. 7. So the final expression f^* annotating the answer $Q(I)$ will be the *sum* of the expressions f_1, f_2 annotating the answers of Q applied to the relations corresponding to $G_{C,1}$ and $G_{C,2}$ respectively.

The relations corresponding to $G_{C,1}$ are

$$R = \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} \begin{matrix} w_1 \\ w_2 \end{matrix} \quad S = \begin{bmatrix} a_1 & c_1 \\ b_1 & c_1 \end{bmatrix} \begin{matrix} v_1 \\ v_2 \end{matrix} \quad T = \begin{bmatrix} c_1 \end{bmatrix} u_1$$

and, the relations corresponding to $G_{C,2}$ are

$$R = \begin{bmatrix} a_2 \end{bmatrix} w_3 \quad S = \begin{bmatrix} a_2 & c_2 \\ a_2 & d_2 \end{bmatrix} \begin{matrix} v_3 \\ v_4 \end{matrix} \quad T = \begin{bmatrix} c_2 \\ d_2 \end{bmatrix} \begin{matrix} u_2 \\ u_3 \end{matrix}$$

Now we focus on the first recursive call at the second level of recursion tree with input co-table subgraph $G_{C,1}$. Note that the table-adjacency graph for this call is the same as G_T . At this level the table decomposition procedure is invoked and the edges of the table-adjacency graph are marked with $+$ and $-$ signs, see Fig. 8. In this figure the common variable set for R, S on the edge (R, S) is $\{x\}$, and for S, T on the edge (S, T) is $\{y\}$. Further, the edge (S, T) is marked with a “ $+$ ” because there are all possible edges between the tuples in S (in this case tuples v_1, v_2) and the tuples in T (in this case u_1). However, tuples in R (here w_1, w_2) and tuples in S (here v_1, v_2) do not have all possible edges between them so the edge (R, S) is marked with a “ $-$ ”.

Table decomposition procedure performs a connected component decomposition using “ $-$ ”-edges, that decomposes G_T in two components $\{R, S\}$ and $\{T\}$. The subset C of common variables collected from the “ $+$ ”-edges across different components will be the variables on the single edge (S, T) , $C = \{y\}$. This variable y is replaced by new *free variables* in all subgoals containing it, which are S and T in our case. So the modified queries for disjoint components returned by the table decomposition procedure are $\widehat{Q}_1() : -R(x)S(x, y_1)$ and $\widehat{Q}_2() : -T(y_2)$. The input graph $G_{C,1}$ is decomposed further into $G_{C,1,1}$ and $G_{C,1,2}$, where $G_{C,1,1}$ will have the edges (w_1, v_1) and (w_2, v_2) , whereas $G_{C,1,2}$ will have no edges and a single vertex u_1 . Moreover, the expression f_1 is the product of f_{11} and f_{12} generated by these two queries respectively. Since the number of tables for \widehat{Q}_2 is only one, and T has a single tuple, by the base step (Step 2) of COMPRO, $f_{12} = u_1$. For expression f_{11} from \widehat{Q}_1 , now the graph $G_{C,1,1}$ can be decomposed using a row decomposition to two connected components with single edges each ((w_1, v_1) and (w_2, v_2) respectively). There will be recursive subcalls on these two components and each one of them will perform a table decomposition (one tuple in every table, so the single edges in both calls will be marked with “ $+$ ”). Hence f_{11} will be evaluated to $f_{11} = w_1 v_1 + w_2 v_2$. So $f_1 = f_{11} \cdot f_{12} = (w_1 v_1 + w_2 v_2) u_1$.

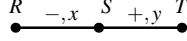


Figure 8: Marked table-adjacency graph for R, S, T .

By a similar analysis as above, it can be shown that the same query Q evaluated on the tables R, S, T given in the above tables give $f_2 = w_3(v_3v_4 + u_2u_3)$. So the overall algorithm is successful and outputs the read-once form $f^* = f_1 + f_2 = (w_1v_1 + w_2v_2)u_1 + w_3(v_3u_2 + v_4u_3)$.

5 Discussion of Time Complexity of Query-Answering Algorithm

Putting together our results from Sections 3 and 4, we propose the following algorithm for answering boolean conjunctive queries without self-joins on tuple-independent probabilistic databases.

Phase 0 (Compute provenance DAG)

Input: query Q , event table rep I

Output: provenance DAG H

Complexity: $O((\frac{ne}{k})^k)$

Phase 1 (Compute co-table graph)

Input: H, Q

Output: table-adjacency graph G_T , co-table graph G_C

Complexity: $O(nm_H + \beta_H m_{co} + k^2 \alpha \log \alpha)$ (Thm. 1)

Phase 2 (Compute read-once form)

Input: event table rep I, Q, G_T, G_C

Output: read-once form f^* or FAIL

Complexity: $O(m_T \alpha \log \alpha + (m_C + n) \min(k, \sqrt{n}))$ (Thm. 2)

Size of the provenance DAG H . Let f be the boolean event expression generated by some query plan for Q on the database I . The number of edges m_H in the DAG H represents the size of the expression f . Since there are exactly k subgoals in the input query Q , one for every table, every prime implicant of f_{IDNF} will have exactly k variables, so the size of f_{IDNF} is at most $\binom{n}{k} \leq (\frac{ne}{k})^k$. Further, the size of the input expression f is maximum when f is already in IDNF. So size of the DAG H is upper bounded by $m_H \leq (\frac{ne}{k})^k$. Again, the “leaves” of the DAG H are exactly the n variables in f . So $m_H \geq n - 1$, where the lower bound is achieved when the DAG H is a tree (every node in H has a unique predecessor); in that case f must be read-once and H is the unique read-once tree of f .

Therefore $n - 1 \leq m_H \leq (\frac{ne}{k})^k$. Although the upper bound is quite high, it is our contention that for practical query plans the size of the provenance graph is much smaller than the size of the

corresponding IDNF.

Data complexity. The complexity dichotomy of [8] is for data complexity, i.e., the size of the query is bounded by a constant. This means that our k and α are $O(1)$. Hence the time complexities of the Phase 1 and Phase 2 are $O(nm_H + \beta_H m_{co})$ and $O(m_C + n) \min(k, \sqrt{n})$ respectively. As discussed above, m_H is $\Omega(n)$ and $O((\frac{ne}{k})^k)$. So one of these two terms may dominate the other based on the relative values of m_H, m_C and m_{co} and of β_H . For example, when $m_H = \theta(n)$, $m_{co} = \theta(m_C) = \theta(n^2)$, and $\beta_H = O(1)$, the first phase takes $O(n^2)$ time, whereas the second phase may take $O(n^{\frac{5}{2}})$ time. However, when $m_H = \Omega(n^{\frac{3}{2}})$, the first phase always dominates.

In any case, we take the same position as [33] that for unsafe [8] queries the competition comes from the approach that does *not* try to detect whether the formulas are read-once and instead uses probabilistic inference [26] which is in general EXPTIME. In contrast, our algorithm runs in PTIME, and works for a larger class of queries than the safe queries [8] (but of course, not on all instances).

Comparisons with other algorithms. For these comparisons we do not restrict ourselves to data complexity, instead taking the various parameters of the problem into consideration.

First consider the **general read-once detection algorithm**. This consists of choosing some plan for the query, computing the answer boolean event expression f , computing its IDNF, and then using the (so far, best) algorithm [12] to check if f is read-once and if so to compute its read-once form. The problem with this approach is that the read-once check is indeed done in time a low polynomial, but *in the size of f_{IDNF}* . For example, consider a boolean query like the one in Example 3. This is a query that admits a plan (the safe plan!) that would generate the event expression $(x_1 + y_1) \cdots (x_n + y_n)$ on an instance in which each R_i has two tuples x_i and y_i . This is a read-once expression easily detected by our algorithm, which avoids the computation of the IDNF.

Next consider the **cograph-help algorithm** that we have already mentioned and justified in Section 2. This consists of our Phase 0 and a slightly modified Phase 1 that computes the co-occurrence graph G_{co} , followed by checking if G_{co} is a cograph using one of the linear-time algorithms given in [6, 19, 3] which also outputs the read-once form if possible. Since Phase 0 and Phase 1 are common we only need to compare the last phases.

The co-graph recognition algorithms will all run in time $O(m_{co} + n)$. Our Phase 2 complexity is better than this when $m_C \min(k, \sqrt{n}) = o(m_{co})$. Although in the worst case this algorithm performs at least as well as our algorithm (since m_C may be $\theta(m_{co})$), (i) almost always the time required in first phases will dominate, so the asymptotic running time of both these algorithms will be comparable, (ii) as we have shown earlier, the ratio $\frac{m_{co}}{m_C}$ can be as large as $\Omega(n^2)$, and the benefit of this could be significantly exploited by caching co-table graphs computed for other queries (see discussions in Section 7), and (iii) these linear time algorithms use complicated data structures, whereas we use simple graphs given as adjacency lists and connectivity-based algorithms, so our algorithms are simpler to implement and may run faster in practice.

Finally we compare our algorithm against that given in [33]. Let us call it the **lineage-tree algorithm** since they take the lineage tree of the result as input as opposed to the provenance DAG as we do. Although [33] does not give a complete running time analysis of the lineage tree

algorithm, for the brief discussion we have, we can make, to the best of our understanding, the following observations.

Every join node in the lineage tree has two children, and every project node can have arbitrary number of children. When the recursive algorithm computes the read-once trees of every child of a project node, every pair of such read-once trees are merged which may take $O(n^2k^2)$ time for every single merge (since the variables in the read-once trees to be merged are repeated). Without counting the time to construct the lineage tree this algorithm may take $O(Nn^2k^2)$ time in the worst case, where N is the number of nodes in the lineage tree.

Since [33] does not discuss constructing the lineage tree we will also ignore our Phase 0. We are left with comparing N with m_H . It is easy to see that the number of edges in the provenance DAG H , $m_H = \theta(N)$, where N is the number of nodes in the lineage tree, when both originate from the same query plan¹¹ Since the lineage-tree algorithm takes $O(Nn^2k^2)$ time in the worst case, and we use $O(nm_H + \beta_H m_{co} + k^2 \alpha \log \alpha) + O((m_C + n) \min(k, \sqrt{n})) = O(nN + \beta_H n^2 + k^2 \alpha \log \alpha + n^{\frac{5}{2}})$. The width β_H of the DAG H in the worst case can be the number of nodes in H . So our algorithm *always* gives an $O(k^2)$ improvement in time complexity over the lineage-tree algorithm given in [33] whereas the benefit can often be more.

6 Related Work

The beautiful complexity dichotomy result of [8] classifying conjunctive queries without self-joins on tuple-independent databases into “safe” and “unsafe” has spurred and intensified interest in probabilistic databases. Some papers have extended the class of safe relational queries [9, 29, 30, 7]. Others have addressed the question of efficient query answering for unsafe queries on *some* probabilistic databases. This includes mixing the intensional and extensional approaches, in effect finding subplans that yield read-once subexpressions in the event expressions [23]. The technique identifies “offending” tuples that violate functional dependencies on which finding safe plans relies and deals with them intensionally. It is not clear that this approach would find the read-once forms that our algorithm finds. The OBDD-based approach in [29] works also for some unsafe queries on some databases. The SPROUT secondary-storage operator [31] can handle efficiently some unsafe queries on databases satisfying certain functional dependencies.

Exactly like us, [33] looks to decide efficiently when the extensional approach is applicable given a conjunctive query without self-joins and a tuple-independent database. We have made comparisons between the two papers in various places, especially in Section 5. Here we only add that that our algorithm deals with different query plans uniformly, while the lineage tree algorithm needs to do more work for non-deep plans. The graph structures used in our approach bear some resemblance to the graph-based synopses for relational selectivity estimation in [34].

The read-once property has been studied for some time, albeit under various names [20, 17, 35, 21, 18, 24, 32]. It was shown [21] that if $RP \neq NP$ then read-once cannot be checked in PTIME for arbitrary monotone boolean formulas, but for formulas in IDNF (as input) read-once can be

¹¹If we “unfold” provenance DAG H to create the lineage tree, the tree will have exactly m_H edges, and the number of nodes in the tree will be $N = m_H + 1$.

checked in PTIME [12]. Our result here sheds new light on another class of formulas for which such an efficient check can be done.

7 Conclusions and Future Work

We have investigated the problem of efficiently deciding when a conjunctive query *without self-joins* applied to a *tuple-independent* probabilistic database representation yields result representations featuring *read-once* boolean event expressions (and, of course, efficiently computing their read-once forms when they exist). We have given a complete and simple to implement algorithm of low polynomial data complexity for this problem, and we have compared our results with those of other approaches.

As explained in the introduction, the results of this paper do not constitute complexity dichotomies. However, there is some hope that the novel proof of completeness that we give for our algorithm may be of help for complexity dichotomy results in the space coordinated by the type of queries and the type of databases we have studied.

Of independent interest may be that we have also implicitly performed a study of an interesting class of monotone boolean formulas, those that can be represented by the provenance graphs of conjunctive queries without self-joins (characterizations of this class of formulas that do not mention the query or the database can be easily given). We have shown that for this class of formulas the read-once property is decidable in low PTIME (the problem for arbitrary formulas is unlikely to be in PTIME, unless $RP=NP$). Along the way we have also given an efficient algorithm for computing the co-occurrence graph of such formulas (in all the other papers we have examined, computing the co-occurrence graph entails an excursion through computing a DNF; this, of course, may be the best one can do for arbitrary formulas, if $RP \neq NP$). It is likely that nicely tractable class of boolean formulas may occur in other database applications, to be discovered.

For further work one obvious direction is to extend our study to larger classes of queries and probabilistic databases [9, 7]. Recall from the discussion in the introduction however, that the class of queries considered should not be able to generate arbitrary monotone boolean expressions. Thus, SPJU queries are too much (but it seems that our approach might be immediately useful in tackling unions of conjunctive queries without self-joins, provided the plans do the unions last).

On the more practical side, work needs to be done to apply our approach to non-boolean queries, i.e., they return actual tables. Essentially, one would work with the provenance graph associated with each table (initial, intermediate, and final) computing simultaneously the co-table graphs of the event expressions on the graph's roots. It is likely that these co-table graphs can be represented together, with ensuing economy.

However we believe that the most practical impact would have the *caching of co-table graphs* at the level of the system, over batches of queries on the same database, since the more expensive step in our algorithm is almost always the computation of the co-table graph (see discussion in Section 5).

This would work as follows, for a fixed database I . When a (let's say boolean for simplicity) conjunctive query Q_1 is processed, consider also the query \bar{Q}_1 which is obtained from Q_1 by replacing each *occurrence* of constants with a distinct fresh FO variable. Moreover if an FO

variable x occurs several times in a subgoal $R_i(\mathbf{x}_i)$ of Q but does *not* occur in any of the other subgoals (i.e., x causes selections but not joins), replace also each occurrence of x with a distinct fresh FO variable. In other words, Q_1 is doing what \bar{Q}_1 is doing, but it first applies some selections on the various tables of I . We can say that \bar{Q}_1 is the “join pattern” behind Q_1 . Next, compute the co-table graph for \bar{Q}_1 on I and *cache* it together with \bar{Q}_1 . It is not hard to see that the co-table graph for Q_1 can be efficiently computed from that of \bar{Q}_1 by a “clean-up” of those parts related to tuples of I that do not satisfy the select conditions of Q_1 .

When another query Q_2 is processed, check if its join-pattern \bar{Q}_2 *matches* any of the join-patterns previously cached (if not, we further cache *its* join-pattern and co-table graph). Let’s say it matches \bar{Q}_1 . Without defining precisely what “matches” means, its salient property is that the co-table graph of \bar{Q}_2 can be efficiently obtained from that of \bar{Q}_1 by another clean-up, just of edges, guided by the table-adjacency graph of \bar{Q}_2 (which is the same as that of Q_2). It can be shown that these clean-up phases add only an $O(n\alpha)$ to the running time.

There are two practical challenges in this approach. The first one is efficiently finding in the cache some join-pattern that matches that of an incoming query. Storing the join-patterns together into some clever data structure might help. The second one is storing largish numbers of cached co-table graphs. Here we observe that they can all be stored with the same set of nodes and each edge would have a list of the co-table graph it which it appears. Even these lists can be large, in fact the number of all possible joint-patterns is exponential in the size of the schema. More ideas are needed and ultimately the viability of this caching technique can only be determined experimentally.

Acknowledgement. We thank Amol Deshpande and Dan Suciu for useful discussions.

References

- [1] L. Antova, T. Jansen, C. Koch, and D. Olteanu. Fast and simple relational processing of uncertain data. In *ICDE*, pages 983–992, 2008.
- [2] O. Benjelloun, A. D. Sarma, A. Y. Halevy, and J. Widom. Uldbs: Databases with uncertainty and lineage. In *VLDB*, pages 953–964, 2006.
- [3] A. Bretscher, D. Corneil, M. Habib, and C. Paul. A simple linear time LexBFS cograph recognition algorithm. *SIAM J. Discrete Math.*, 22(4):1277–1296, 2008.
- [4] R. Cavallo and M. Pittarelli. The theory of probabilistic databases. In *VLDB*, pages 71–81, 1987.
- [5] D. G. Corneil, H. Lerchs, and L. S. Burlingham. Complement reducible graphs. *Discrete Appl. Math.*, 3(3):163–174, 1981.

- [6] D. G. Corneil, Y. Perl, and L. K. Stewart. A linear recognition algorithm for cographs. *SIAM J. Comput.*, 14(4):926–934, 1985.
- [7] N. N. Dalvi, K. Schnaitter, and D. Suciu. Computing query probability with incidence algebras. In *PODS*, pages 203–214, 2010.
- [8] N. N. Dalvi and D. Suciu. Efficient query evaluation on probabilistic databases. In *VLDB*, pages 864–875, 2004.
- [9] N. N. Dalvi and D. Suciu. The dichotomy of conjunctive queries on probabilistic structures. In *PODS*, pages 293–302, 2007.
- [10] N. N. Dalvi and D. Suciu. Management of probabilistic data: foundations and challenges. In *PODS*, pages 1–12, 2007.
- [11] N. Fuhr and T. Rölleke. A probabilistic relational algebra for the integration of information retrieval and database systems. *ACM Trans. Inf. Syst.*, 15(1):32–66, 1997.
- [12] M. C. Golumbic, A. Mintz, and U. Rotics. Factoring and recognition of read-once functions using cographs and normality and the readability of functions associated with partial k -trees. *Discrete Appl. Math.*, 154(10):1465–1477, 2006.
- [13] E. Grädel, Y. Gurevich, and C. Hirsch. The complexity of query reliability. In *PODS*, pages 227–234, 1998.
- [14] T. J. Green, G. Karvounarakis, Z. G. Ives, and V. Tannen. Update exchange with mappings and provenance. In *VLDB*, pages 675–686, 2007.
- [15] T. J. Green, G. Karvounarakis, and V. Tannen. Provenance semirings. In *PODS*, pages 31–40, 2007.
- [16] T. J. Green and V. Tannen. Models for incomplete and probabilistic information. In *EDBT Workshops*, pages 278–296, 2006.
- [17] V. A. Gurvich. Repetition-free Boolean functions. *Uspehi Mat. Nauk*, 32(1(193)):183–184, 1977.
- [18] V. A. Gurvich. Criteria for repetition-freeness of functions in the algebra of logic. *Soviet Math. Dokl.*, 43(3):721–726, 1991.
- [19] M. Habib and C. Paul. A simple linear time algorithm for cograph recognition. *Discrete Applied Mathematics*, 145(2):183 – 197, 2005.
- [20] J. P. Hayes. The fanout structure of switching functions. *J. Assoc. Comput. Mach.*, 22(4):551–571, 1975.
- [21] L. Hellerstein and M. Karpinski. Learning read-once formulas using membership queries. In *COLT*, pages 146–161, 1989.

- [22] T. Imielinski and W. L. Jr. Incomplete information in relational databases. *J. ACM*, 31(4):761–791, 1984.
- [23] A. Jha, D. Olteanu, and D. Suciu. Bridging the gap between intensional and extensional query evaluation in probabilistic databases. In *EDBT*, pages 323–334, 2010.
- [24] M. Karchmer, N. Linial, I. Newman, M. E. Saks, and A. Wigderson. Combinatorial characterization of read-once formulae. *Discrete Mathematics*, 114(1-3):275–282, 1993.
- [25] G. Karvounarakis, Z. G. Ives, and V. Tannen. Querying data provenance. In *SIGMOD*, pages 951–962, 2010.
- [26] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, Cambridge, MA, 2009.
- [27] L. V. S. Lakshmanan, N. Leone, R. B. Ross, and V. S. Subrahmanian. Probview: A flexible probabilistic database system. *ACM Trans. Database Syst.*, 22(3), 1997.
- [28] I. Newman. On read-once boolean functions. In M. S. Paterson, editor, *Boolean Function Complexity*, pages 25–34. Cambridge University Press, 1992.
- [29] D. Olteanu and J. Huang. Using obdds for efficient query evaluation on probabilistic databases. In *SUM*, pages 326–340, 2008.
- [30] D. Olteanu and J. Huang. Secondary-storage confidence computation for conjunctive queries with inequalities. In *SIGMOD*, pages 389–402, 2009.
- [31] D. Olteanu, J. Huang, and C. Koch. Sprout: Lazy vs. eager query plans for tuple-independent probabilistic databases. In *ICDE*, pages 640–651, 2009.
- [32] J. Peer and R. Pinter. Minimal decomposition of boolean functions using non-repeating literal trees. In *IFIP Workshop on Logic and Architecture Synthesis*, 1995.
- [33] P. Sen, A. Deshpande, and L. Getoor. Read-once functions and query evaluation in probabilistic databases. In *To appear in VLDB*, 2010.
- [34] J. Spiegel and N. Polyzotis. Graph-based synopses for relational selectivity estimation. In *SIGMOD Conference*, pages 205–216, 2006.
- [35] L. G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, 1984.
- [36] E. Zimányi. Query evaluation in probabilistic relational databases. *Theor. Comput. Sci.*, 171(1-2):179–219, 1997.

A Generalization of Example 1

Example 5.

$R =$	a_1	x_1	$T =$	c_1	u_1
	b_1	y_1		c_2	u_2
	a_2	x_2		d_2	v_2
	\dots	\dots		\dots	\dots
	a_{2n-1}	x_{2n-1}		c_{2n-1}	u_{2n-1}
	b_{2n-1}	y_{2n-1}		c_{2n}	u_{2n}
	a_{2n}	x_{2n}		d_{2n}	v_{2n}

$S =$	a_1	c_1	z_1
	b_1	c_1	z_2
	a_2	c_2	z_3
	a_2	d_2	z_4
	\dots	\dots	\dots
	a_{2n-1}	c_{2n-1}	z_{4n-3}
	b_{2n-1}	c_{2n-1}	z_{4n-2}
	a_{2n}	c_{2n}	z_{4n-1}
	a_{2n}	d_{2n}	z_{4n}

Answer event expression with standard plan:

$$x_1 z_1 u_1 + y_1 z_2 u_1 + \dots + x_{2n} z_{4n-1} u_{2n} + x_{2n} z_{4n} v_{2n} \quad (5)$$

Equivalent to:

$$(x_1 z_1 + y_1 z_2) u_1 + \dots + x_{2n} (z_{4n-1} u_{2n} + z_{4n} v_{2n}) \quad (6)$$

which is in read-once form and whose probability can be computed in time $O(n)$.

B PTIME Probability Computation for Non-Read-Once Expressions

The following example shows that there exists a query Q and a probabilistic database instance D such that the expression E for evaluation of query Q on database D is not read-once but still the probability of E being true can be computed in poly-time.

Example 6. The database D has three tables $R(A), S(A, B), T(B)$. Table S has n tuples. The tuple in the $2i - 1$ -th row has tuple (a_i, b_i) and the tuple in the $2i$ -th row has tuple (a_{i+1}, b_i) . We suppose that S is a deterministic relation, i.e. all the tuples in S belong to S with probability one and are annotated with true. If $n = 2k$ then table R has $k + 1$ tuples, if $n = 2k - 1$ then R has k tuples. The tuple in row j of R is a_j and is annotated by x_{2j-1} . Table T has k tuples, where $n = 2k - 1$ or $n = 2k$: the tuple in row j is b_j and is annotated by x_{2j} . It can be verified that the expression E_n annotating the answer to the query $Q() = R(A), S(A, B), T(B)$ is

$$E_n = x_1 x_2 + x_2 x_3 + \dots x_{n-1} x_n + x_n x_{n+1}.$$

An example with $n = 2$ is given in Figure 9. It can be easily verified that for all $n > 1$, the expression E_n is not read-once.

$$S = \begin{array}{|c|c|} \hline a_1 & x_1 \\ \hline a_2 & x_3 \\ \hline \end{array} \quad R = \begin{array}{|c|c|c|} \hline a_1 & b_1 & z_1 = 1_{\mathbb{B}} \\ \hline a_2 & b_1 & z_2 = 1_{\mathbb{B}} \\ \hline a_2 & b_2 & z_3 = 1_{\mathbb{B}} \\ \hline \end{array} \quad T = \begin{array}{|c|c|} \hline b_1 & x_2 \\ \hline b_2 & x_4 \\ \hline \end{array}$$

Figure 9: Illustration with $n = 3$, $E_3 = x_1x_2 + x_2x_3 + x_3x_4$.

Next we show that $P_n = P(E_n)$ can be computed in poly-time in n by dynamic programming. Note that P_1 can be computed in $O(1)$ time. Suppose for all $\ell < n$, P_ℓ is computed and stored in an array. Then

$$\begin{aligned} P_n &= P(x_1x_2 + \dots + x_{n-2}x_{n-1} + x_{n-1}x_n + x_nx_{n+1}) \\ &= P(x_1x_2 + \dots + x_{n-2}x_{n-1} + x_{n-1}x_n) + P(x_nx_{n+1}) \\ &\quad - P(x_nx_{n+1}[x_1x_2 + \dots + x_{n-2}x_{n-1} + x_{n-1}x_n]) \\ &= P_{n-1} + P(x_nx_{n+1}) \end{aligned} \tag{7}$$

$$- P(x_nx_{n+1}[x_1x_2 + \dots + x_{n-2}x_{n-1} + x_{n-1}x_n]) \tag{8}$$

Observe that:

$$\begin{aligned} &P(x_nx_{n+1}[x_1x_2 + \dots + x_{n-2}x_{n-1} + x_{n-1}x_n]) \\ &= P(x_{n+1})P(x_n[x_1x_2 + \dots + x_{n-2}x_{n-1} + x_{n-1}x_n]) \\ &= P(x_{n+1})P(x_1x_2x_n + \dots + x_{n-2}x_{n-1}x_n + x_{n-1}x_n) \text{ (idempotency)} \\ &= P(x_{n+1})P(x_1x_2x_n + \dots + x_{n-3}x_{n-2}x_n + x_{n-1}x_n) \text{ (absorption)} \\ &= P(x_{n+1})P(x_n)P(x_1x_2 + \dots + x_{n-3}x_{n-2} + x_{n-1}) \\ &= P(x_nx_{n+1})[P(x_1x_2 + \dots + x_{n-3}x_{n-2}) + P(x_{n-1})] \\ &= P(x_nx_{n+1})[P_{n-3} + P(x_{n-1})] \end{aligned} \tag{9}$$

From (8) and (9),

$$\begin{aligned} P_n &= P_{n-1} + P(x_nx_{n+1}) - P(x_nx_{n+1})[P_{n-3} + P(x_{n-1})] \\ &= P_{n-1} + P(x_nx_{n+1})[1 - P_{n-3} - P(x_{n-1})] \end{aligned}$$

Since the variables x_i -s are independent, $P(x_nx_{n+1}) = P(x_n)P(x_{n+1})$, and while computing P_n , P_{n-1} and P_{n-3} are already available. Hence P_n can be computed in linear time.

C Proofs from Section 3

Proof of Lemma 4.

LEMMA 4. *Algorithm COMPCoTABLE adds an edge (x, y) to G_C if and only if x, y together appear in some implicant in f_{IDNF} and the tables containing x, y are adjacent in G_T .*

Proof. Suppose two variables x, y belong to the same implicant in f_{IDNF} , and their tables are adjacent in G_T . Then by Lemma 2, there is a $--$ -node $u \in \text{lca}(x, y)$, and $x \in \text{Var}(v_1), y \in \text{Var}(v_2)$ for two distinct successors v_1, v_2 of u . When the algorithm processes the node u , if an edge between x, y is not added in a previous step, the edge will be added. This shows the completeness of algorithm COMPCoTABLE.

Now we show the soundness of the algorithm. Consider two variables x, y such that either the tables containing them are not adjacent in G_T or they do not belong together in any of the implicants in f_{IDNF} . If the tables containing x, y are not adjacent in G_T , clearly, the algorithm never adds an edge between them – so let us consider the case when x, y do not belong to the same implicant in f_{IDNF} . Then by Lemma 2, there is no $--$ -node $u \in \text{lca}(x, y)$.

Consider any iteration of the algorithm and consider that a node u is processed by the algorithm in this iteration. If u is a $+$ -node or if either $x \notin \text{Var}(u)$ or $y \notin \text{Var}(u)$, again no edge is added between x, y . So assume that, u is a $--$ -node and $x, y \in \text{Var}(u)$. Then u is a common ancestor of x and y . But since $u \notin \text{lca}(x, y)$, by definition of least common ancestor set, there is a successor v of u such that v is an ancestor of both x, y and therefore, $x, y \in \text{Var}(v)$. However, by Corollary 3, since x or y cannot belong to two distinct successors of node u , node v must be the unique successor of u such that $x, y \in \text{Var}(v)$. Since COMPCoTABLE only joins variables from two distinct children, no edge will be added between x and y in G_C . \square

C.1 Time Complexity of COMPCoTABLE

First we prove the following two lemmas bounding the number of times any given pair of variables x, y are considered by the algorithm. The first lemma shows that the variables x, y are considered by algorithm COMPCoTABLE to add an edge between them in G_{co} only when they together appear in an implicant in f_{IDNF} , i.e. only if the edge actually should exist in G_{co} .

Lemma 8. *Consider any two variables x, y and a $--$ -node u . If x, y do not appear together in an implicant in f_{IDNF} , x, y do not belong to the variable sets $\text{Var}(v_1), \text{Var}(v_2)$ for two distinct successors v_1, v_2 of u .*

Proof. This easily follows from Lemma 2 which says that if x, y do not appear together in an implicant in f_{IDNF} , then there is no $--$ -node in $\text{lca}(x, y)$. So for every $--$ -node u , either (i) one of x and $y \notin \text{Var}(u)$, or, (ii) there is a unique successor v of u which is a common ancestor of x, y , i.e. both $x, y \in \text{Var}(v)$ (uniqueness follows from Corollary 3). \square

The second lemma bounds the number of times a pair x, y is considered by the algorithm to add an edge between them.

Lemma 9. Suppose $x, y \in \text{Var}(f)$ be such that they together appear in an implicant f_{IDNF} . Then algorithm COMPCOTABLE considers x, y in Step 10 to add an edge between them maximum β_H times, where β_H is the width of the provenance DAG H .

Proof. Note that the check in Step 10 is performed only when the current node u is a \cdot -node. Consider any \cdot -node u . (i) if either x or y is not in $\text{Var}(u)$, clearly, x, y are not checked in this step, otherwise, (ii) if both $x, y \in \text{Var}(u)$, and $x, y \in \text{Var}(v)$ for a unique child v of u , then also x, y are not checked at this step, otherwise, (iii) if u joins x, y , i.e., $x \in \text{Var}(v_1), y \in \text{Var}(v_2)$ for two distinct children v_1, v_2 of u , then only x, y are considered by the algorithm in Step 10. (and after this node u is processed, both x, y appear in $\text{Var}(u)$).

However, since the query does not have any self-joins, the only time two variables x, y appear in two distinct successors of a \cdot -node u when the query plan joins a subset of tables containing the table for x with a subset of tables containing the table for y . So the pair x, y is multiplied at a unique layer of H , and the total number of times they are multiplied cannot exceed the total number of nodes in the layer which is at most the width β_H of the DAG H . \square

Now we complete the running time analysis of algorithm COMPCOTABLE.

Lemma 10. Given the table-adjacency graph G_T and input query plan H , algorithm COMPCOTABLE can be implemented in time $O(\beta_H m_{co} + nm_H)$ time, where m_{co} is the number of edges in the co-occurrence graph, m_H is the number of edges in the DAG H , β_H is the width of the DAG H and $n = |\text{Var}(f)|$.

Proof. Initialization step can be done in $O(n)$ time. The topological sort can be done in $O(m_H + |V(H)|)$ time by any standard algorithm.

At every node $u \in V(H)$, to compute set $\text{Var}(u)$, the algorithm scans $O(d_u)$ successors of u , where d_u = the outdegree of node u in H . Although by Corollary 3, for every two distinct children v_1, v_2 of a \cdot -node u , $\text{Var}(v_1) \cap \text{Var}(v_2) = \emptyset$, they may have some overlap when u is a $+$ -node, and here the algorithm incurs an $O(nm_H)$ cost total as follows: (i) create an n -length boolean array for u initialized to all zero, (ii) scan $\text{Var}(v)$ list of every successor v of u , for a variable $x \in \text{Var}(v)$, if the entry for x in the boolean array is false mark it as true, (iii) finally scan the boolean array again to collect the variables marked as true for variables in $\text{Var}(u)$. At every node $u \in V(H)$, the algorithm spends $O(nd_u)$ time, where d_u = the outdegree of node u in H . Hence the total time across all nodes = $\sum_{u \in V(H)} O(nd_u) = O(nm_H)$.

Every check in Step 10, i.e., whether an edge (x, y) has already been added and whether the tables containing x, y are adjacent in G_T can be done in $O(1)$ time using $O(n^2 + k^2) = O(n^2)$ space. Further, by Lemma 8 and 9, the number of such checks performed is $O(\beta_H m_{co})$. Since $\text{Var}(f) \subseteq V(H)$, and H is connected, $n \leq |V(H)| \leq |E(H)|$. Hence the total time complexity is $O(nm_H + \beta_H m_{co})$. \square

We can now finish to prove Theorem 1. As shown in Section 3.2, computation of the table-adjacency graph G_T takes $O(k^2 \alpha \log \alpha)$ time and this proves the second part of Theorem 1. The time complexity analysis in Lemma 10 also holds when we modify COMPCOTABLE to compute the co-occurrence graph G_{co} instead of the co-table graph G_C : the only change is that we do not check whether the tables containing x, y are adjacent in G_T . Further, we do not need to precompute the graph G_T . This proves the first part and completes the proof of Theorem 1.

D Proofs from Section 4

Modified query in Algorithm 4 evaluates the same expression:

Lemma 11. *Suppose $I = R_{i_1}[T'_{i_1}], \dots, R_{i_p}[T'_{i_p}]$ be the set of input tables to the table decomposition procedure TD and let $Q() : -R_{i_1}(\mathbf{x}_{i_1}), \dots, R_{i_p}(\mathbf{x}_{i_p})$ be the input query. Then the expression g generated by evaluating query Q on I is exactly the same as evaluating \widehat{Q} on I , where $\widehat{Q} = \widehat{Q}_1, \dots, \widehat{Q}_\ell$ is the conjunction of modified queries \widehat{Q}_j returned by the procedure TD for groups $j = 1$ to ℓ .*

Proof. We prove that a set of p tuple variables taken from p tables satisfy the original input query Q' if and only if they satisfy the modified query \widehat{Q} . Since the new query subgoals make some of the original variables *free*, by replacing them with new variables, clearly, if a set of tuples satisfy the original query they also satisfy the modified query. So we prove that the modified query does not introduce any erroneous collection of tuples in the final answer.

Consider a set of tuple variables s.t. the corresponding tuples satisfy the modified query. Let us partition these variables according to the ℓ groups of tables as computed by procedure TD. Consider component j of the partition and any table R_i in component j . Recall that C_i is the set of all variables on the “+”-edges having one end at the table i in component j . A “+” edge between table R_i and $R_{i'}$ implies that the edges between every tuple in R_i and every tuple in $R_{i'}$ exist, which in turn implies that, all tuples in R_i and $R_{i'}$ must have the *same* values of the attributes corresponding to $C_e = \mathbf{x}_i \cap \mathbf{x}_{i'}$. Then any set of p tuples taken from p tables must have the same value of attributes corresponding to variables in C_e . In other words, every variable $z \in C_i$ can be replaced by a new free variable z^j in every table R_i in component j (note that $C_i \subseteq \mathbf{x}_i$) without changing the final solution. \square

Proof of Lemma 5.

LEMMA 5. *At any step of the recursion, if row decomposition is successful then table decomposition is unsuccessful and vice versa.*

Proof. Consider any step of the recursive procedure, where the input tables are $R_{i_1}[T'_{i_1}], \dots, R_{i_q}[T'_{i_q}]$ ($\forall j, T'_{i_j} \subseteq T_{i_j}$), input query is $Q'() : -R_{i_1}(\mathbf{x}_{i_1}), \dots, R_{i_q}(\mathbf{x}_{i_q})$, and the induced subgraphs of G_C and G_T on current sets of tuples and tables are G'_C and G'_T respectively.

Suppose row decomposition is successful, i.e., it is possible to decompose the tuples in G'_C into $\ell \geq 2$ connected components. Consider any two tables R_i, R_j such that the edge (R_i, R_j) exists in G'_T , and consider their sub-tables $R_i[T_i^1]$ and $R_j[T_j^2]$ taken from two different connected components in G'_C . Consider two arbitrary tuples $x \in T_i^1$ and $x' \in T_j^2$. Since x and x' belong to two different connected components in G'_C , then there is no edge (x, x') in G'_C . Hence by Step 3 of the table decomposition procedure, this edge (R_i, R_j) will be marked by “−”. Since (R_i, R_j) was an arbitrary edge in G'_T , all edges in G'_T will be marked by “−” and there will be a unique component in G'_T using “−” edges. Therefore, the table decomposition procedure will be unsuccessful.

Now suppose table decomposition is successful, i.e., G'_T can be decomposed into $\ell \geq 2$ components using “−” edges. Note that wlog. we can assume that the initial table-adjacency graph

G_T is connected. Otherwise, we can run the algorithm on different components of G_T and multiplied the final expressions from different components at the end. Since the procedure TD returns induced subgraphs for every connected components, the input subgraph G'_T is always a connected graph at every step of the recursion. Now consider any two tables R_i and R_j from two different groups components such that (R_i, R_j) edge exists in G'_T (such a pair must exist since the graph G'_T is connected). Since this edge is between two components of a successful table decomposition procedure, it must be marked with “+”. This implies that for any tuple $x \in R_i$ and any tuple $x' \in R_j$, the edge (x, x') exists in G'_C (which follows from Step 3 of this procedure). This in turn implies that row decomposition must fail at this step since the tables R_i, R_j cannot be decomposed into two disjoint components and the graph G'_C will be connected through these tuples. \square

Proof of Lemma 6.

LEMMA 6. (Soundness) *If the algorithm returns with success, then the expression f^* returned by the algorithm COMPRO is equivalent to the expression $Q(I)$ generated by evaluation of query Q on instance I . Further, the output expression f^* is in read-once form.*

Proof. We prove the lemma by induction on n , where $n = \bigcup_{i=1}^k |T_i|$. The base case follows when $n = 1$. In this case there is only one tuple x , hence k must be 1 as well, and therefore, the algorithm returns x in Step 2. Here the algorithm trivially returns with success and outputs a read-once form. The output is also correct, since computation of co-table graph ensures that there is no unused tuple in the tables, and the unique tuple x is the answer to query Q on database I .

Suppose the induction hypothesis holds for all databases with number of tuples $\leq n - 1$ and consider a database with n tuples. If $k = 1$, then irrespective of the query, all tuples in table R_1 satisfies the query $Q() : -R_1(\mathbf{x}_1)$ (again, there are no unused tuples), and therefore the algorithm correctly returns $\sum_{x \in T_1} x$ as the answer which is also in read-once form. So let us consider the case when $k \geq 2$.

(1) Suppose the current recursive call successfully performs row decomposition, and $\ell \geq 2$ components $\langle R_1[T_1^1], \dots, R_k[T_k^1] \rangle, \dots, \langle R_1[T_1^\ell], \dots, R_k[T_k^\ell] \rangle$ are returned. By the row decomposition algorithm, it follows that for $x \in T_i^j$ and $x' \in T_{i'}^{j'}$, x and x' do not appear together in any monomial in the DNF equivalent for $Q(I)$. So the tuples which row decomposition puts in different components do not join with each other and then the final answer of the query is the union of the answers of the queries on the different components. Then the final expression is the sum of the final expressions corresponding to the different components. Since all components have $< n$ tuples and the algorithm did not return with error in any of the recursive calls, by the inductive hypothesis all the expressions returned by the recursive calls are the correct expressions and are in read-once form. Moreover these expressions clearly do not share variables – they correspond to tuples from different tables since the query does not have a self-join. We conclude that the final expression computed by the algorithm is the correct one and is in read-once form.

(2) Otherwise, suppose the current step successfully performs table decomposition. Let $\ell \geq 2$ groups $\mathbf{R}_1, \dots, \mathbf{R}_\ell$ are returned. Correctness of table decomposition procedure, i.e., correctness of the expression $f^* = f_1 \dots f_\ell$, when all the recursive calls return successfully follows from Lemma 11 using the induction hypothesis (the algorithm multiplies the expressions returned by

different groups which themselves are correct by the inductive hypothesis). Further, since all components have $< n$ tuples, and the algorithm did not return with error in any of the recursive calls, all expressions returned by the recursive calls are in read-once form. Since they do not share any common variable, the final output expression is also in read-once form. \square

Proof of Lemma 7.

LEMMA 7. (**Completeness**) *If the expression $Q(I)$ is read-once, then the algorithm COMPRO returns the unique read-once form f^* of the expression.*

Proof. Suppose the expression is read-once and consider the tree representation T^* of the unique read-once form f^* of the expression (T^* is in *canonical form* and has alternate levels of $+$ and \cdot nodes, which implies that every node in T^* must have at least two children.). We prove the lemma by induction on the height h of tree T^* .

First consider the base case. If $h = 1$, then the tree must have a single node for a single tuple variable x . Then k must be 1 and the algorithm returns the correct answer. So consider $h \geq 2$.

(1) Consider the case when root of the tree is a $+$ node. If $h = 2$, since we do not allow union operation, k must be 1 and all the tuples must belong to the same table R_1 . This is taken care of by Step 2 of COMPRO. If $h > 2$, then k must be ≥ 2 and the answer to the join operation must be non-empty. Every child of the root node corresponds to a set of monomials which will be generated by the equivalent DNF expression f_{DNF} for the subtree rooted at that child. Note that no two variables in two different children of the root node can belong to any monomial together since the tree T^* is in read-once form. In other words, they do not share an edge in G_C . Hence the component formed by the set of variables at a child will not have any edge to the set of variables at another child of the root node. This shows that all variables at different children of the root node will belong to different components by the row decomposition procedure.

Now we show that variables at different children of the root node are put to different components by the row decomposition procedure, which shows that the row decomposition algorithm will divide the tuples *exactly* the same was as the root of T^* divides tuples among its children. Since T^* is in canonical read-once form and has alternate levels of $+$ and \cdot nodes, then row decomposition cannot be done within the same subtree of a $+$ node. So all variables in a subtree must form a connected component. Since the root has ≥ 2 children, in this case we will have a successful row decomposition operation. By inductive hypothesis, since the subtrees rooted at the children of the root are all in read-once form, the recursive calls of the algorithm on the corresponding subtrees are successful. Hence the overall algorithm at the top-most level will be successful.

(2) Now consider the case when root of the tree is a \cdot node. Note that the \cdot operator can only appear as a result of join operation. If the root has $\ell' \geq 2$ children $c_1, \dots, c_{\ell'}$, then every tuple x in the subtree at c_j joins with every tuple y in the subtree at $c_{j'}$ for every pair $1 \leq j \neq j' \leq \ell'$. Moreover since the query does not have a self join, x, y must belong to two different tables, which implies that there is an edge (x, y) in G_C between every pair of tuples x, y from subtrees at $c_j, c_{j'}$ respectively. Again, since we do not allow self-join, and T^* is in read-once form, the tuples in the subtrees at $c_j, c_{j'}$ must belong to different tables if $j \neq j'$. In other words, the tables R_1, \dots, R_k are partitioned into ℓ' disjoint groups $\mathbf{R}'_1, \dots, \mathbf{R}'_{\ell'}$.

Next we argue that $\ell = \ell'$ and the partition returned by the table decomposition procedure $\mathbf{R}_1, \dots, \mathbf{R}_\ell$ is identical to $\mathbf{R}'_1, \dots, \mathbf{R}'_{\ell'}$ upto a permutation of indices. Consider any pair \mathbf{R}'_j and $\mathbf{R}'_{j'}$. Since the tuple variables in these two groups are connected by a \cdot operator, all tuples in all tables in \mathbf{R}'_j join with all tuples in all tables in $\mathbf{R}'_{j'}$. In other words, for any pair of tuples x, x' from $R_{i_1} \in \mathbf{R}'_j$ and $R_{i_2} \in \mathbf{R}'_{j'}$, there is an edge (x, x') in co-occurrence graph G_{co} . Hence if there is a common subset of join attributes between R_{i_1} and R_{i_2} , i.e. the edge (R_{i_1}, R_{i_2}) exists in G_T , it will be marked by a “+” (all possible edges between tuples will exist in the co-table graph G_T). So the table decomposition procedure will put \mathbf{R}'_j and $\mathbf{R}'_{j'}$ in two different components. This shows that $\ell \geq \ell'$. However, since T^* is in read-once form and has alternate levels of $+$ and \cdot nodes, no \mathbf{R}'_j can be decomposed further using join operation (i.e. using “+” marked edges by the table decomposition procedure); therefore, $\ell' = \ell$. Hence our table decomposition operations exactly outputs the groups $\mathbf{R}'_1, \dots, \mathbf{R}'_{\ell'}$. By the inductive hypothesis the algorithm returns with success in all recursive calls, and since $\ell' = \ell \geq 2$, the table decomposition returns with success. So the algorithm returns with success. \square

D.1 Time Complexity of COMPRO

Here we discuss the time complexity of algorithm COMPRO in detail and show that algorithm COMPRO runs in time $O(m_T \alpha \log \alpha + (m_C + n) \min(k, \sqrt{n}))$. We divide the time complexity computation in two parts: (i) total time required to compute the modified queries across *all* table decomposition steps performed by the algorithm (this will give $O(m_T \alpha \log \alpha)$ time) and (ii) total time required for all other steps: here we will *ignore* the time complexity for the modified query computation step and will get a bound of $(m_C + n) \min(k, \sqrt{n})$. First we bound the time complexity of individual row decomposition and table decomposition steps.

Lemma 12. *The row decomposition procedure as given in Algorithm 3 runs in time $O(m'_C + n')$, where $n' = \sum_{j=1}^q |T_{i_j}|$ = the total number of input tuples to the procedure, and m'_C = the number of edges in the induced subgraph of G_C on these n' tuples.*

Proof. The row decomposition procedure only runs a connectivity algorithm like BFS/DFS to compute the connected components. Then it collects and returns the tuples and computes the induced subgraphs in these components. All these can be done in linear time in the size of the input graph which is $O(m'_C + n')$. \square

Next we show that the table decomposition can be executed in time $O(m'_C + n')$ as well.

Lemma 13. *The table decomposition procedure as given in Algorithm 4 runs in time $O(m'_C + n')$, ignoring the time required to compute the modified queries $\widehat{Q_j}$ where $n' = \sum_{j=1}^q |T'_{i_j}|$ = the total number of input tuples to the procedure, and m'_C = the number of edges in the induced subgraph of G_C on these n' tuples.*

Proof. Step 3 in the table decomposition procedure marks edges in G'_T using G'_C . Let us assume that G'_C has been represented in a standard adjacency list. Consider a table $R_j[T'_j]$, where $T'_j \subseteq T_j$ and let d be the degree of R_j in G'_T . Now a linear scan over the edges in G'_C can partition the edges

e from a tuple $x \in T'_j$ in table R_j into E_1, \dots, E_d , where E_q ($q \in [1, d]$) contains all edges from x to tuples x' , belonging to the q -th neighbor of R_j . A second linear scan on these grouped adjacency lists computed in the previous step is sufficient to mark every edge in G'_T with a “+” or a “−”: for every neighbor q of R_j , say $R_{j'}$, for every tuple x in T'_j , scan the q -th group in adjacency list to check if x has edges with all tuples in $R_{j'}[T'_{j'}]$. If yes, then all tuples in $R_{j'}$ also have edges to all tuples in R_j , and the edge $(R_j, R_{j'})$ is marked with a “+”. Otherwise, the edge is marked with a “−”. Hence the above two steps take $O(m'_C + n' + m'_T + k')$ time, where k' and m'_t are the number of vertices (number of input tables) and edges in the subgraph G'_T .

Finally returning the induced subgraphs of G'_T for the connected components and decomposition of the tuples takes $O(m'_C + n' + m'_T + k')$ time. Since $n' \geq k'$ and $m'_C \geq m'_T$, not considering the time needed to recompute the queries, step, the total time complexity is bounded by $O(m'_C + n')$. \square

The next lemma bounds the total time required to compute the modified queries over all calls to the recursive algorithm.

Lemma 14. *The modified queries \widehat{Q}_j over all steps can be computed in time $O(m_T \alpha \log \alpha)$, where α is the maximum size of a subgoal.*

Proof. We will use a simple charging argument to prove this lemma. For an edge $e = (R_i, R_j)$ in G_T , the common variable set $C_e = \mathbf{x}_i \cap \mathbf{x}_j$ ¹² can be computed by (i) first sorting the variables in $\mathbf{x}_i, \mathbf{x}_j$ in some fixed order, and then (ii) doing a linear scan on these sorted lists to compute the common variables. Here we to compute the set C_e . Hence this step takes $O(\alpha \log \alpha)$ time. Alternatively, we can use a hash table to store the variables in \mathbf{x}_i , and then by a single scan of variables in \mathbf{x}_j and using this hash table we can compute the common attribute set C_e in $O(\alpha)$ expected time. When C_e has been computed in a fixed sorted order for every edge e incident on R_i to a different component, the lists C_e -s can be repeatedly merged to compute the variables set $C_i = \bigcup_e C_e$ in $O(d_i \alpha)$ time (note that even after merging any number of C_e sets, the individual lists length are bounded by the subgoal size of R_i which is bounded by α). However, instead of considering the total time $O(d_i \alpha)$ for the node R_i in G_T , we will *charge* every such edge $e = (R_i, R_j)$ in G_T for this merging procedure an amount of $O(\alpha)$. So every edge e from R_i to an R_j in different component gets a charge of $O(\alpha \log \alpha)$.

Suppose we charge the outgoing edges (R_i, R_j) from R_i to different components by a fixed cost of P , $P = O(\alpha)$ in the above process. From the table decomposition procedure it follows that, the common join attributes are computed, and the query is updated, only when the edge (R_i, R_j) belongs to the *cut* between two connected components formed by the “−” edges. These edges then get *deleted* by the table decomposition procedure: all the following recursive calls consider the edges *inside* these connected components and the edges *between* two connected components are never considered later. So each edge in the graph G_T can be *charged* at most once for computation of common join attributes and this gives $O(m_C) \alpha \log \alpha$ as the total time required for this process.

Finally, the variables in \mathbf{x}_i can also be replaced by new variables using the sorted list for C_i in $O(\alpha)$ time, so the total time needed is $O(m_C \alpha \log \alpha + n \alpha) = O(m_C \alpha \log \alpha)$ (since we assumed G_T for the query Q is connected without loss of generality). \square

¹²We abuse the notation and consider the *sets* corresponding to *vectors* $\mathbf{x}_i, \mathbf{x}_j$ to compute C_e

Now we show that the depth of the recursion tree is $O(\min(k, \sqrt{n}))$ and in every level of the tree, the total time required is at most $O(m_C + n)$. Let us consider the recursion tree of the algorithm COMPRO and wlog. assume that the top-most level performs a row decomposition. Since the size of the table-adjacency subgraph G'_T is always dominated by the co-table subgraph G'_C at any recursive call of the algorithm, we express the time complexity of the algorithm with k tables, and, n tuples and m edges in the subgraph G'_C as $T_1(n, m, k)$, where the top-most operation is a row decomposition. Further, every component after row decomposition has exactly k tables and therefore must have at least k tuples, because, we assumed wlog. that initial table adjacency graph G_T is connected and there is no unused tuples in the tables. Similarly, $T_2(n, m, k)$ denotes the time complexity when the top-most operation is a table decomposition operation. Note that at every step, for row decomposition, every tuple and every edge in G'_C goes to exactly one of the recursive calls of the algorithm; however, the number of tables k remains unchanged. On the other hand, for table decomposition operation, every tuple goes to exactly one recursive call, every edge goes to at most one such calls (edges between connected components are discarded), and every table goes to exactly one call. Recall that the row and table decomposition alternates at every step, and the time required for both steps is $O(m + n)$ (not considering computation of modified queries at every table decomposition steps) so we have the following recursive formula for $T_1(n, m, k)$ and $T_2(n, m, k)$.

$$\begin{aligned}
T_1(n, m, k) &= O(m + n) + \sum_{j=1}^{\ell} T_2(n_j, m_j, k) \\
&\quad \text{where } \sum_{j=1}^{\ell} n_j = n, \sum_{j=1}^{\ell} m_j = m, n_j \geq k \forall j \\
T_2(n, m, k) &= O(m + n) + \sum_{j=1}^{\ell} T_1(n_j, m_j, k_j) \\
&\quad \text{where } \sum_{j=1}^{\ell} n_j = n, \sum_{j=1}^{\ell} m_j \leq m, \sum_{j=1}^{\ell} k_j = k
\end{aligned}$$

where n_j, m_j and k_j are the total number of tuples and edges in G'_C , and the number of tables for the j -th recursive call (for row decomposition, $k_j = k$). For the base case, we have $T_2(n_j, m_j, 1) = O(n_j)$ – for $k = 1$, to compute the the read once form, $O(n_j)$ time is needed; also in this case $m_j = 0$ (a row decomposition cannot be a leaf in the recursion tree for a successful completion of the algorithm). Moreover, it is important to note that for a successful row or table decomposition, $\ell \geq 2$.

If we draw the recursion tree for $T_1(n, m_C, k)$ (assuming the top-most operation is a row-decomposition operation), at every level of the tree we pay cost at most $O(m_C + n)$. This is because the tuples and edges go to at most one of the recursive calls and k does not play a role at any node of the recursion tree (and is absorbed by the term $O(m_C + n)$).

Now we give a bound on the height of the recursion tree.

Lemma 15. *The height of the recursion tree is upper bounded by $O(\min(k, \sqrt{n}))$.*

Proof. Every internal node has at least two children and there are at most k leaves (we return from a path in the recursion tree when k becomes 1). Therefore, there are $O(k)$ nodes in the tree and the height of the tree is bounded by $O(k)$ (note that both the number of nodes and the height may be $\Theta(k)$ when the tree is not balanced).

Next we show that the height of the recursion tree is also bounded by $4\sqrt{n}$. The recursion tree has alternate layers of table and row decomposition. We focus on only the table decomposition layers, the height of the tree will be at most twice the number of these layers. Now consider any arbitrary path P in the recursion tree from the root to a leaf where the number of table decompositions on P is h . Suppose that in the calls $T_2(n, m, k)$, the values of n and k along this path (for the table decomposition layers) are $(n_0, k_0), (n_1, k_1), \dots, (n_h, k_h)$, where $k_0 = k$ and $n_0 \leq n$ (if the top-most level has a table-decomposition operation, then $n_0 = n$). We show that $h \leq 2\sqrt{n}$.

Let's assume the contradiction that $h > 2\sqrt{n}$ and let's look at the first $p = 2\sqrt{n}$ levels along the path P . If at any j -th layer, $j \in [1, p]$, $k_j \leq 2\sqrt{n} - j$, then the number of table decomposition steps along P is at most $2\sqrt{n}$: every node in the recursion tree has at least two children, so the value of k decreases by at least 1. The number of table-decomposition layers after the j -th node is at most k_j , and the number of table-decomposition layers before the j -th node is exactly j . Therefore, the total number of table-decomposition layers is $\leq 2\sqrt{n}$.

Otherwise, for all $j \in [1, p]$, $k_j > 2\sqrt{n} - j$. Note that $n_j \leq n_{j-1} - k_{j-1}$: there is a row decomposition step between two table decompositions, and every component in the j -th row decomposition step will have at least k_j nodes. If this is the case, we show that $n_p < 0$. However,

$$\begin{aligned}
n_p &\leq n_{p-1} - k_{p-1} \\
&\leq n_{p-2} - k_{p-2} - k_{p-1} \\
&\vdots \\
&\leq n_0 - \sum_{j=0}^{p-1} k_j \\
&\leq n - \sum_{j=0}^{p-1} k_j \\
&\leq n - \sum_{j=0}^{2\sqrt{n}-1} (2\sqrt{n} - j) \\
&= n - \sum_{j=1}^{2\sqrt{n}} j \\
&= n - \frac{2\sqrt{n}(2\sqrt{n} + 1)}{2} \\
&= n - 2n - \sqrt{n} \\
&< 0
\end{aligned}$$

which is a contradiction since n_p is the number of nodes at a recursive call and cannot be negative. This shows that along any path from root to leaves, the number of table decomposition layers is

bounded by $2\sqrt{n}$ which in turn shows that the height of the tree is bounded by $4\sqrt{n}$. \square

Since total time needed at every step of the recursion tree is $O(m_C + n)$, we have the following corollary,

Corollary 1. *Not considering the time complexity to compute the modified queries by the table decomposition procedure, the algorithm COMPRO runs in time $O((m_C + n) \min(k, \sqrt{n}))$.*

The above corollary together with Lemma 14 (which says that to compute the modified queries $O(m_T \alpha \log \alpha)$ time suffices) shows that COMPRO runs in time $O(m_T \alpha \log \alpha + (m_C + n) \min(k, \sqrt{n}))$.